Etudes biophysiques et structurales du complexe de réplication des Rhabdoviridae: La phosphoprotéine et ses interactions avec la nucléoprotéine

Cedric Leyrat

Ecole Doctorale Chimie Sciences du Vivant

THESE

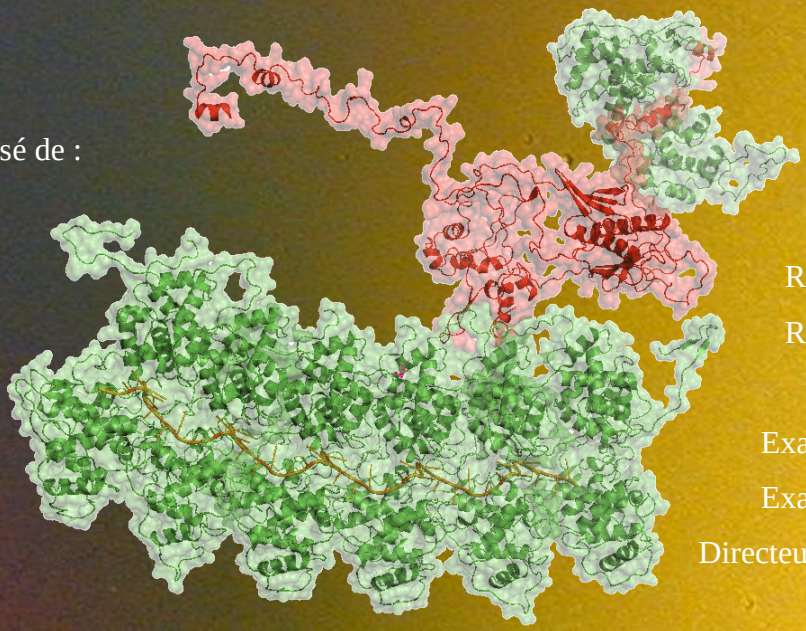
Pour l'obtention du diplôme de
Docteur de l'Université Joseph Fourier – Grenoble I

Spécialité : Chimie – Biologie

Présentée et soutenue publiquement par
Cédric LEYRAT

Le vendredi 19 novembre 2010

Etudes biophysiques et structurales du complexe de réplication des *Rhabdoviridae*: La phosphoprotéine et ses interactions avec la nucléoprotéine



Devant le jury composé de :

Dr. Noël Tordo	Rapporteur
Dr. Sonia Longhi	Rapporteur
Pr. Dan Kolakofsky	Examinateur
Pr. Valdur Saks	Examinateur
Pr. Marc Jamin	Directeur de thèse

Thèse réalisée à l'**UVHCI** (Unit of Virus Host Cell Interactions UMI 3265 UJF-EMBL-CNRS)
D'octobre 2007 à septembre 2010

Résumé

Le virus de la stomatite vésiculaire (VSV) et le virus de la rage (RV) appartiennent à la famille des *Rhabdoviridae* et sont les représentants prototypiques respectifs des genres *Vesiculovirus* et *Lyssavirus*. Ces virus sont enveloppés, et leur génome est composé d'une molécule d'ARN de polarité négative qui encode successivement cinq protéines virales dont la nucléoprotéine (N) et la phosphoprotéine (P), qui forment, avec l'ARN polymérase ARN dépendante virale (L), le complexe de transcription et de réplication des *Rhabdoviridae*. L'objectif de mon travail de thèse consistait à obtenir des informations structurales concernant la phosphoprotéine et plusieurs de ses fragments, à l'état libre en solution, ou en complexe avec N, et ce afin de mieux comprendre les mécanismes impliqués dans la multiplication virale. La protéine P interagit avec N en formant deux types de complexes distincts : le complexe N-ARN-P, résultant de l'association de P avec les complexes N-ARN, qui constituent la matrice utilisée pour la transcription et la réplication virale, et le complexe N^0-P, qui correspond à la fixation par P d'une molécule de nucléoprotéine N^0 monomérique, empêchant ainsi la fixation non spécifique de N à l'ARN et permettant l'encapsidation des génomes et anti-génomes néo-synthétisés lors de la réplication virale. Les travaux réalisés ont permis de mettre en évidence la nature modulaire des dimères de phosphoprotéine, et de caractériser la région N-terminale intrinsèquement désordonnée de P (P_{NTD}) impliquée dans la fixation à N^0. Nous avons montré que le domaine C-terminal impliqué dans la fixation à N-ARN (P_{CTD}) constituait un module structural indépendant, et modélisé à l'échelle atomique la structure multi-domaine partiellement désordonnée de P en solution. Nous avons également construit un modèle atomique *in silico* de l'interaction entre N-ARN et P, validé expérimentalement, et enfin résolus par diffraction des rayons X la structure du complexe N^0-P_{NTD} à l'état cristallin. Les travaux présentés concernant la structure de la phosphoprotéine libre ou en complexe avec la nucléoprotéine permettent de mieux comprendre le rôle de P dans le complexe de réplication, et soulignent l'importance du désordre moléculaire et des changements conformationnels dans les processus de reconnaissance moléculaire impliquant N et P. Les modèles structuraux des complexes N-ARN-P et N^0-P constituent par ailleurs des nouvelles cibles intéressantes pour le développement de composés antiviraux.

Abstract

Vesicular Stomatitis Virus (VSV) and Rabies Virus (RV) belong to the *Rhabdoviridae* family and are prototypic members of the genus *Vesiculovirus* and *Lyssavirus*, respectively. VSV and RV are envelopped, single stranded, negative sense RNA viruses, and their genome encode successively for five viral proteins including nucleoprotein (N) and phosphoprotein (P), which form, together with the viral RNA dependent RNA polymerase (L), the *Rhabdoviridae* transcription and replication complex. The aim of this work was to obtain structural information regarding the structure of phosphoprotein and of several of its fragments, free or in complex with N, in order to better understand the mechanisms involved in viral replication. The P protein interacts with N by forming two distinct types of complexes : the N-ARN-P complex results from the association of P with the N-RNA complex, which constitutes the active template for transcription and replication ; the N^0-P complex arises from the attachment of P to a N^0 molecule, thus preventing binding to non specific RNA and allowing for the specific encapsidation of newly synthesized genomes and antigenomes during viral replication. The data presented here reveals the modular organization of P dimers and highlights the intrinsically disordered character of its N-terminal region involved in N^0 binding (P_{NTD}). It is shown that the C-terminal domain of P (P_{CTD}) is an autonomous folding unit that can bind to N-RNA complexes. The atomic scale modeling of P in solution shows that the protein behaves as a partially disordered, highly flexible multidomain protein. Additionally, we constructed a structural model of the interaction between N-RNA and P *in silico*, which was validated experimentally, and we solved the crystal structure of the N^0-P_{NTD} complex using X-Ray diffraction. The work presented here concerning the structure of P, free or in complex with N enhances our understanding of the role of P in the viral replication complex, and emphasizes the importance of structural disorder and conformational changes in the molecular recognition processes involving N and P. The structural models for the N-ARN-P and N^0-P complexes also represent attractive new targets for the development of antiviral drugs.

Remerciements

La vie est un jeu de hasard dans lequel on est amené à faire des choix dont la portée nous dépasse souvent. J'étais loin d'imaginer, lors de ma rencontre avec les Pr. Marc JAMIN et Pr. Rob RUIGROK il y a 4 ans environ, la succession d'évènements qui m'amènerait jusqu'à l'écriture de cette page. Cette rencontre, qui faisait suite à un premier contact avec le monde de la virologie, grâce au module de Virologie Humaine des Pr. Rob RUIGROK et Pr. Patrice MORAND, un module proposé en Master 1 à l'Université Joseph Fourier de Grenoble. Merci à eux de m'avoir fait découvrir les virus. Ma fascination pour les virus, considérés en tant qu'assemblages de macromolécules, probablement inoculée par le Pr. Rob RUIGROK, s'est ajoutée à la fascination pour les molécules (déjà présente à mon arrivée) et a grandement contribué à me faire apprécier doublement ces 4 années passés à l'UVHCI.

Je ne pourrai jamais suffisamment remercier le Pr. Marc JAMIN pour avoir été un Directeur de thèse à l'esprit ouvert et débordant d'idées, pour avoir toujours été disponible, m'avoir impliqué dans divers projets très intéressants, m'avoir judicieusement conseillé et orienté en fonction des problèmes qui se présentaient, et m'avoir toujours soutenu dans la réalisation de mes idées.

Je remercie le Dr. Noël TORDO et le Dr. Sonia LONGHI d'avoir accepté d'être les rapporteurs de ce travail. Je suis aussi extrêmement reconnaissant envers le Pr. Dan KOLAKOFSKY et le Pr. Valdur SAKS pour leur participation à mon jury de thèse.

Je remercie le directeur de l'EMBL Grenoble et directeur de l'UVHCI, le Dr. Stephen CUSACK de m'avoir accueilli au sein de son institut, où règne un environnement scientifique unique grâce au PSB regroupant l'EMBL, l'UVHCI, l'ILL, l'ESRF, et l'IBS.

Je tiens également à remercier nos collaborateurs :

- Les Dr. Malene Ringkjøbing JENSEN et Dr. Martin BLACKLEDGE du laboratoire de RMN de l'Institut de Biologie Structurale à Grenoble.

- Dr. Danielle BLONDEL au laboratoire de Virologie Moléculaire et Structurale à Gif-sur-Yvette pour les échanges de plasmides mutés et les tests de double hybride.

- Dr. Noël TORDO, et Guillaume CASTEL de l'Institut Pasteur à Paris (collaboration projet ANR, ANRAGE),

- Dr. Giuseppe ZACCAI à l'Institut Laue Langevin à Grenoble (SANS).

Merci au Dr. Nicolas TARBOURIECH, de nous avoir enseigné l'art de la cristallisation des protéines, à Filip et à moi, de nous avoir fait profiter de son expérience et de son expertise dans le maniement de la baguette magique magnétique, de nous avoir emmené sur les lignes de Xtalo, et de m'avoir efficacement coaché sur la résolution de la structure.

Un grand merci au « Maître », le Dr. Euripedes de Almeida RIBEIRO Jr, qui a totalement « boosté » les projets de recherche initiés au laboratoire et qui a pu me guider pendant ces 4 dernières années en m'enseignant notamment à utiliser les pouvoirs de la force pour « sentir » la concentration des protéines dans la solution.

Un grand merci également à Filip YABUKARSKI pour sa motivation, sa détermination, son endurance et sa contribution au projet N^0-P.

Je tiens à remercier sincèrement Francine GERARD pour m'avoir coaché à mon arrivée au laboratoire, et aussi bien entendu pour toutes ses contributions aux projets.

Je voudrais remercier nos microscopistes électroniques, le Dr. Irina GUSTCHE, le Dr. Guy SCHOEHN, le Dr. Leandro ESTROZI et en particulier Ambroise DESFOSSES, qui a réalisé avec moi nombre d'expériences

biophysiques bizarres concernant la fibrillation de peptides induite par la cuisson, et aussi bien sûr pour m'avoir enseigné les bases du scripting en C-shell./*{}#...

Je tiens à remercier chaleureusement Louiza ZERRAD pour son support irréprochable lors de toutes les mesures SAXS sur ID14-3.

Je tiens à remercier Julien PERARD, Grand Gourou de la biologie moléculaire, pour la collaboration sur le projet de modélisation de cette molécule d'ARN géante *a.k.a* HCV IRES, mais aussi et surtout pour m'avoir fait partager une petite partie de son savoir concernant toutes les techniques de laboratoire : tes conseils avisés ont contribués à la réussite d'un grand nombre de mes expériences.

Un grand merci au Père MARTINEZ pour sa patience et sa bonne humeur et au Père IVANOV pour son humour Bulgare et sa grande gentillesse.

Je tiens à remercier tous les membres du laboratoire, anciens ou actuels, merci à vous tous.

Je tiens à remercier plus particulièrement Emilie DUFOUR, Aurélie ALBERTINI, Charlotte SUEUR, Céleste SELE, Eric THIERRY, Lucy FREEMAN, Alex DIAS, Julien LUPO, Pierre-Alain COLY *a.k.a* le PAC, Audrey, Laurence, Mohammed HABIB, Frank, Patricia, Charles, Nolwenn, Sylvie, Juan, Pascal, Grégory, Majida, Marlyse, Thibault, Sophie, Xavier, Denis, Steffie, Yan, Anton, Estelle, Delphine, Adam, Eve, Django Reinhardt, Jimi Hendrix, Bob Marley, Michael Jackson, la civilisation Maya…et bien d'autres…

Sans oublier ceux que l'on appelle « the people from the underground » : David, Ganesh, Martinellours, Pauline, Miriam, Aurélien, Emilie, et les autres…

Je tiens à remercier le personnel de l'EMBL et de l'UVHCI qui nous rend la vie si simple : merci à Marie-Laure, Alice, Annie, Jocelyne, Virginie, Françoise, Sandrine, Christelle, Pierre, Lionel, José, Benjamin, Aymeric, Mikaël, Franck, Denis…

Je voudrais aussi dire merci à ma famille et mes amis, qui m'ont toujours soutenu. Merci à vous de m'avoir encouragé même si les protéines et les virus, ça ne vous parle pas plus que ça. Merci à mes parents, qui m'ont laissé le choix de faire des études, de longues études et qui m'ont toujours aidé dans mes démarches. Merci à vous deux du fond du cœur.

Enfin, merci de tout coeur à Axelle, ma compagne, qui a fait preuve d'une grande patience et m'a toujours encouragé et soutenu d'une manière irréprochable pendant ces 3 ans de thèse …

A mes parents, Jean et Jeanne,
A Isabelle et Sire Marcus,
A Axou.

Table des matières

Abréviations

2D : deux dimensionnel
3D : trois dimensionnel
Å : Angstrom (10^{-10}m)
aa : acide aminé
ABL : *Australian Bat Lyssavirus*
ADN : Acide DésoxyriboNucléique
ARN : Acide RiboNucléique
ARNm : ARN messager
CD: Dichroïsme Circulaire
CK-II: Caséine Kinase II
CTD : C-Terminal Domain
CVS: *Challenge Virus Strain*
Da / kDa : Dalton / kiloDalton
DLS : Dynamic Light Scattering
EBL: *European Bat Lyssavirus*
E. coli: *Escherichia coli*
EDTA : *Ethylene Diamine Tetraacetic Acid*
ESRF : *European Synchroton Radiation Facility*
fs: femtoseconde
G: Glycoprotéine
gcq: GROMACS cool quote
GdnHCl: Guanidinium hydrochloride
h : heure
HSQC : Heteronuclear Single Quantum Coherence
IBS : Institut de Biologie Structurale
ICTV : *International Committee on Taxonomy of Viruses*
IDP : Intrinsically Disordered Protein
IDR : Intrinsically Disordered Region
kb : kilobase
L: ARN polymérase ARN dépendante virale
M / mM: Molaire / milliMolaire
M: protéine Matrice
MALLS : *Multi Angle Laser Light Scattering*
MD : Dynamique Moléculaire
MM: Masse Moléculaire
nm : nanomètre.
ns : nanoseconde
nt : nucléotide
NTD : N-Terminal Domain
N : Nucléoprotéine
NC: Nucéocapside
OMS : Organisation Mondiale de la Sante
P : Phosphoprotéine
P_{CED} : Domaine Central de la phosphoprotéine
P_{CTD} : Domaine C-terminal de la phosphoprotéine
P_{NTD} : Région N-terminal de la phosphoprotéine
PAGE : *PolyAcrylamide Gel Electrophoresis*

PCR: *Polymerase Chain Reaction*
PDB : *Protein Data Bank*
PEG : *Polyethylene Glycol*
pH : potentiel Hydrogène
pI : point Isoélectrique
ps : picoseconde
RER : reticulum endoplasmique rugueux
R_g : Rayon de giration
R_S : Rayon de Stokes
RI : réfractométrie
RMN: Résonance Magnétique Nucléaire
RNP: RiboNucleoProteine ou RiboNucleocapside
RMSD: *root mean square deviation*
RMSF: *root mean square fluctuation*
rpm : tour par minute
RSV : virus Respiratoire Syncitial humain
RT-PCR : Reverse Transcriptase PCR
RV : virus de la rage
SANS: *Small Angle Neutron Scattering*
SAXS: *Small Angle X-ray Scattering*
SDS: Sodium Docécyle Sulfate
SEC: *Size Exclusion Chromatography*
SPR : résonance plasmonique de surface (*surface plasmon resonance*)
SNC: Système Nerveux Central
T : Température
TMAO : Triméthylamine N-oxyde
UV : Ultraviolet
VSV : Virus de la Stomatite Vésiculaire

CHAPITRE I: INTRODUCTION

"A serious and good philosophical work could be written consisting entirely of jokes."

(Ludwig Wittgenstein, Austrian Philosopher)

I. Introduction générale

A. Historique & Généralités

Les premières observations des symptômes de la rage consignées par écrit proviennent du code Eshnunna, rédigé en Mésopotamie au XXIIIème siècle avant J.C. (Théodoridès 1986):

« Si un chien est fou et si les autorités ont porté ce fait à la connaissance de son propriétaire ; si ce dernier ne le garde pas chez lui, s'il mord un homme et provoque sa mort, il devra payer deux tiers d'un mina (40 shekels) d'argent. S'il mord un esclave et provoque sa mort, il paiera 15 shekels d'argent. »

Les premières vaccinations antirabiques furent pratiquées par Pierre Galtier sur des ruminants dès 1881 (Galtier 1881), et la première vaccination antirabique chez l'homme a été pratiquée il y a 125 ans en 1885 par Louis Pasteur (Pasteur 1886). Cependant, cette pathologie reste aujourd'hui un problème sérieux de santé publique dans de nombreux pays du monde et particulièrement dans les pays pauvres.

En 1881, l'équipe de Louis Pasteur montre que le principal site de réplication du virus rabique est le système nerveux central et qu'il est plus facile de transmettre la rage par inoculation intracérébrale de substances virulentes. Ces découvertes lui permettent, en réalisant plusieurs passages en série du virus par inoculation intracérébrale sur le lapin, d'obtenir une souche virale fixe qui après atténuation par dessiccation sera utilisée en 1884 dans la mise au point d'un protocole d'immunisation sur des chiens (Steele and Baer 1975). Pasteur l'essaie avec succès, pour la première fois, le 6 juillet 1885, sur un enfant de 9 ans, Joseph Meister, mordu 14 fois par un chien enragé et dont la mort semblait inévitable. Un an plus tard, Pasteur rapporte des résultats satisfaisants de traitements humains, réalisés après exposition, et propose alors la création d'un centre de vaccination : l'Institut Pasteur.

Comme pour bien d'autres maladies dans le passé, les méthodes de protection contre la rage ont devancé la connaissance exacte de l'agent causal. Aujourd'hui, si la nature virale de la maladie ne fait plus aucun doute, à l'époque de Pasteur, la nature de l'agent infectieux faisait l'objet de nombreuses controverses. Les premiers « agents filtrables » ont été identifiés en 1892 par Dmitri Iwanowski qui étudiait le virus de la mosaïque du tabac. Ces travaux ont été poursuivis par Martinus Beijerinck, qui en 1898 définit pour la première fois le « virus » comme une entité distincte (Beijerinck 1898). Pourtant, en 1903, Aldechi Négri qui a mis en évidence des corps d'inclusion - qui portent son nom - dans les préparations de tissus nerveux contaminés par la rage, croyait que l'agent responsable était un micro-organisme parasite de la famille des protozoaires (Negri 1903). Pasteur parlait d'un microbe minuscule différent des bactéries ordinaires et Babès pensait que des bactéries renfermaient l'agent causal sous forme d'un corps minuscule (Babes 1912). Les travaux du XXème siècle ont permis des avancées importantes sur la connaissance de l'agent rabique, de sa structure et du diagnostic de la

maladie. Sellers propose, en 1927, une méthode rapide de détection des corps de Négri (Young and Sellers 1927) et Webster et Clow sont les premiers, en 1936, à propager le virus sur culture cellulaire, permettant la mise au point d'un nouveau modèle d'étude expérimental. Ils ouvrent aussi la voie de la production d'un nouveau type de vaccin (Webster and Clow 1936).

Ce n'est pourtant qu'en 1957 que Lwoff donne une définition conceptuelle claire du virus (Lwoff 1957):

- le virion ne possède qu'un seul type d'acide nucléique : soit ADN, soit ARN.

- il se reproduit uniquement à partir de son matériel génétique.

- il est incapable de croître et de se diviser seul.

- son génome ne contient aucune information permettant la synthèse d'enzymes impliquées dans le métabolisme intermédiaire.

- c'est un parasite absolu de la cellule hôte.

Il faut attendre les années 1960 pour que le virus soit enfin observé en microscopie électronique (Atanasiu, Lepine et al. 1963) et 25 ans supplémentaires pour que la biologie moléculaire commence à élucider la structure de son patrimoine génétique (Tordo, Poch et al. 1986; Tordo, Poch et al. 1986) et à en appréhender la diversité (Bourhy, Kissi et al. 1993). Aujourd'hui, la découverte de certaines familles de virus telles que les *Retroviridae* capables d'utiliser à la fois l'ADN et l'ARN ou encore la caractérisation des Mimivirus dont le génome encode 911 protéines (La Scola, Audic et al. 2003; Raoult, Audic et al. 2004; Renesto, Abergel et al. 2006), ont partiellement remis en cause certaines des affirmations de Lwoff, mais les virus restent cependant des parasites absolus de la cellule hôte.

Même si l'histoire est riche en découvertes et avancées majeures, force est de constater que d'immenses progrès restent à faire. En matière de santé publique d'abord puisque la maladie perdure dans une grande partie du monde malgré l'existence d'un vaccin efficace. L'Organisation Mondiale de la Santé estime à 55000 le nombre de décès humains imputables à la rage chaque année dans le monde (WHO Fact sheet No.99), mais ces chiffres sont probablement très en dessous de la réalité (Cleaveland, Fevre et al. 2002; Knobel, Cleaveland et al. 2005). Même si l'incidence de la rage, humaine et animale, est bien plus importante dans les pays en voie de développement (Asie, Afrique), tous les pays peuvent être touchés par des cas sporadiques autochtones ou par des cas importés. Outre l'aspect humain, cette pathologie a aussi des répercussions économiques importantes : coût des traitements antirabiques post-exposition, pertes de bétail considérables (plusieurs dizaines de milliers de bovins chaque année en Amérique latine). En recherche fondamentale, les étapes clefs de la biologie virale demeurent largement incomprises, tant au niveau des mécanismes d'expression du virus qu'à celui de son neurotropisme, ou des dysfonctionnements qui provoquent la mort.

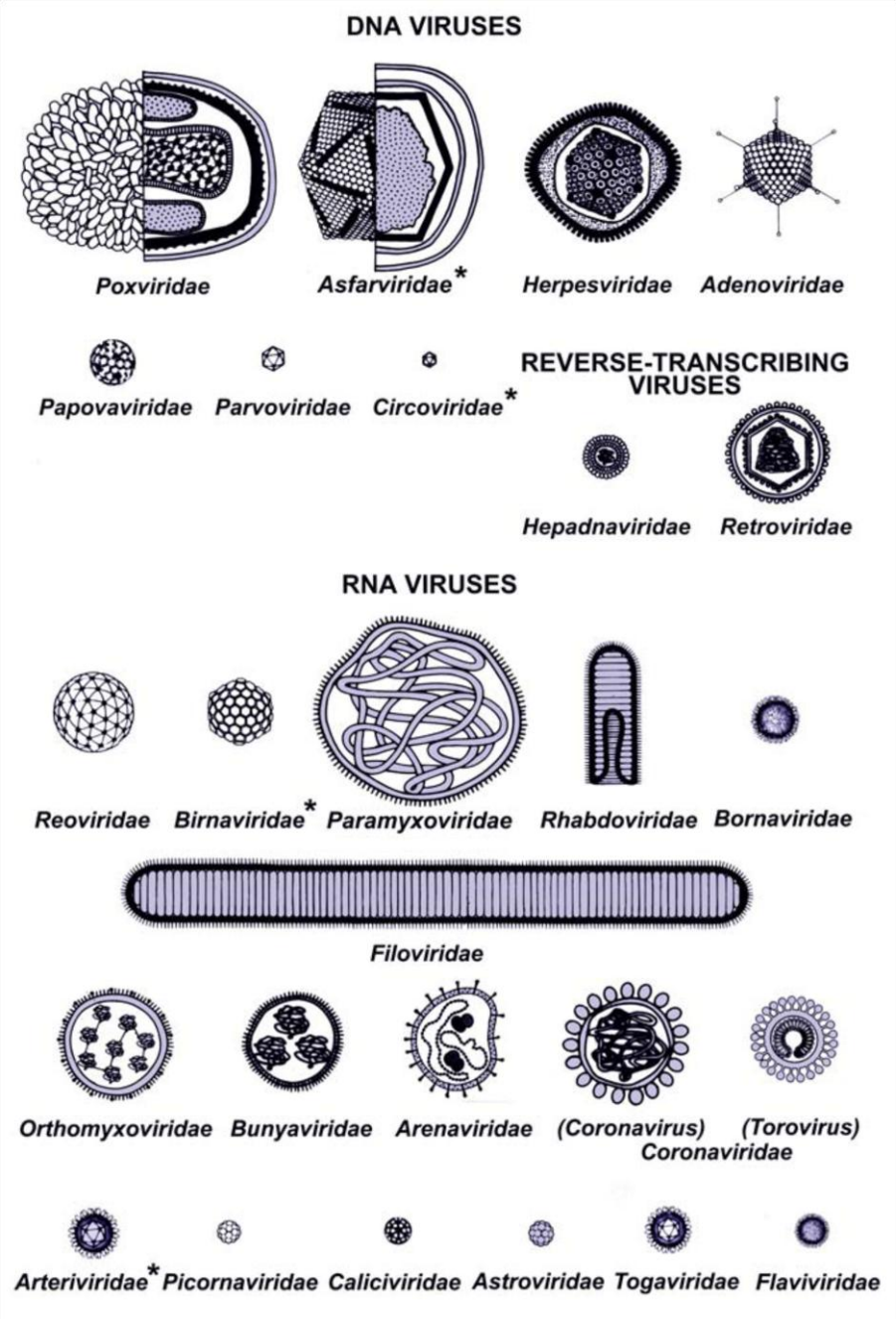


Figure 1: Morphologie et dimensions relatives de différentes familles de virus comprenant des pathogènes humains ou animaux. Les virions sont dessinés à l'échelle, le génome ainsi qu'une coupe interne de leur structure est représentée pour certains, à l'exception des virions les plus petits pour lesquels seul la taille et la symétrie apparaissent. Les virus issus de familles marquées d'un astérisque n'infectent généralement pas l'homme (Büchen-Osmond 2003).

De nos jours, les progrès dans le domaine de la biologie moléculaire et de la génétique, ainsi que l'apparition de la génétique inverse, ont permis un développement considérable de la recherche en virologie. L'étude des virus a été grandement facilitée par le développement de la cristallographie aux rayons X, avec l'étude de cristaux de protéines virales et de virus entiers, qui ont permis de mieux comprendre l'organisation et la fonction des protéines présentes au sein de la particule virale. On dénombre à l'heure actuelle plus de 4 000 virus, capables d'infecter des organismes appartenant à tous les domaines du vivant, des

Archaebactéries aux Eucaryotes. Ils possèdent des génomes (ADN monocaténaire ou bicaténaire, circulaires ou linéaires, ARN monocaténaire (polarité négative ou positive) ou bicaténaire, linéaire ou segmenté) (classification de Baltimore) et des morphologies variées, qui prennent en compte la symétrie de la capside ou de la nucléocapside (icosaédrique par exemple pour l'*adenovirus*, ou hélicoïdale pour les nucléocapsides des virus à ARN négatif), et la présence ou l'absence d'enveloppe (Figure 1). Toutefois, il apparaît clairement aujourd'hui que la compréhension des processus de réplication requiert d'intégrer des approches biologiques, biochimiques, biophysiques, bioinformatiques et structurales. La notion de virus est à peine vieille d'un siècle et la virologie est un domaine de recherche qui suscite énormément d'intérêt dans la communauté scientifique, en raison de la variabilité des virus d'un point de vue structural, génomique et pathologique. L'étude des virus constitue également un outil pour comprendre les mécanismes moléculaires de la cellule.

Les familles de virus représentées dans la Figure 1 incluent de nombreux pathogènes humains responsables de problèmes majeurs de santé publique. Les plus célèbres comprennent le virus de la grippe (*Orthomyxoviridae*), le virus respiratoire syncytial humain et le virus de la rougeole (*Paramyxoviridae*), le virus Ebola (*Filoviridae*), les *rotavirus* mais aussi les *adenovirus*, les *astrovirus* et les *enterovirus* qui provoquent des maladies diarrhéiques, ainsi que le virus de l'immunodéficience humaine (*Retroviridae*). Alors que des vaccins efficaces existent contre plusieurs de ces virus, de nombreuses personnes, et plus particulièrement des enfants, continuent de mourir chaque année, par manque d'accès aux traitements préventifs ou curatifs. Aussi, on sait maintenant que certains cancers sont associés à la présence de virus, comme le cancer du col de l'utérus, qui est la deuxième cause de mortalité par cancer dans le monde chez la femme (environ 230 000 décès et 500 000 nouveaux cas par an), et, qui dans plus de 95% des cas est lié à la persistance de *papillomavirus* (*Papovaviridae*), le lymphome de Burkitt associé au virus d'Epstein-Barr, ou encore le cancer du foie lié à la présence des virus des hépatites B (*Hepadnaviridae*) et C (*Flaviviridae*).

B. Les Virus à ARN négatif

Family/genus	Genome size (kb)	Type virus[a]	Hosts[b]
Mononegavirales (nonsegmented)			
Rhabdoviridae	11-15		
Vesiculovirus		VSV Indiana	vertebrates/insects
Lyssavirus		Rabies virus	vertebrates
Ephemerovirus		BEFV	vertebrates
Novirhabdovirus		IHNV	fish
Cytorhabdovirus		LNYV	plants
Nucleorhabdovirus		PYDV	plants
Filoviridae	19		
Marburg-like viruses		Marburg virus	vertebrates

Ebola-like viruses		Zaire Ebola virus	vertebrates
Paramyxoviridae	15-20		
Respirovirus		Sendai	vertebrates
Morbillivirus		Measles	vertebrates
Rubulavirus		Mumps	vertebrates
Henipavirus		Hendra	vertebrates
Avulavirus		NDV	birds
Pneumovirus		HRSV	vertebrates
Metapneumovirus		TRTV	Turkeys
Bornaviridae	9		
Bornavirus		BDV	vertebrates
Segmented negative-strand RNA viruses			
Orthomyxoviridae	13-15 in 6-8 segments		
Influenzavirus A		Influenza A	vertebrates
Influenzavirus B		Influenza B	vertebrates
Influenzavirus C		Influenza C	vertebrates
Thogoto-like viruses		Thogoto	vertebrates/insects
Bunyaviridae	11-20 in 3 segments		
Orthobunyavirus		Bunyamwera	vertebrates
Hantavirus		Hantaan	vertebrates
Nairovirus		Dugbe	vertebrates
Phlebovirus		Rift Valley fever	vertebrates
Tospovirus		TSWV	Plants
Arenaviridae	10-14 in 2 segments		
Arenavirus		LCMV	vertebrates

[a] *Abbreviations:* VSV Indiana, vesicular stomatitis virus Indiana serotype; BEFV, bovine ephemeral fever virus; IHNV, infectious hematopoietic necrosis virus; LNYV, lettuce necrotic yellows virus; PYDV, potato yellow dwarf virus; NDV, Newcastle disease virus; HRSV, human respiratory syncytial virus; TRTV, turkey rhinotracheitis virus; BDV, Borna disease virus; TSWV, tomato spotted wilt virus; LCMV, lymphocytic choriomeningitis virus.
[b] In all cases, vertebrates include humans as susceptible organisms.

Tableau 1: Classification des différents virus appartenant à l'ordre des M*ononegavirales*.

Les virus à ARN négatif sont des virus enveloppés, leur membrane dérivant de la membrane de la cellule hôte infectée, dont l'ARN viral génomique simple brin de polarité négative est complémentaire de l'ARN messager. Les virus à ARN négatif ont des morphologies variables (Figure 2) et infectent un large spectre d'hôtes allant des plantes aux invertébrés et aux mammifères. Ces virus possèdent un génome variant de 9 à 19 kb, codant entre 5 et 11 protéines seulement (Tableau 1) et ils sont à l'origine de nombreuses pathologies humaines. En effet, certains sont responsables d'infections respiratoires aigües (virus de la grippe, virus respiratoire syncytial humain), d'encéphalites graves (virus de la rage, virus Nipah), ou de fièvres hémorragiques (virus Ebola, Lassa, de la fièvre de Crimée Congo). On distingue les virus dont le génome est segmenté et ceux dont le génome est non-segmenté. Les premiers, divisés en trois familles, ont un génome constitué de plusieurs molécules ou « segments » d'ARN simple brin de polarité négative :

- *Arenaviridae* (2 segments, virus Lassa)
- *Bunyaviridae* (3 segments, virus de la fièvre de la vallée du Rift)
- *Orthomyxoviridae* (7 à 8 segments, virus de la grippe).

Les virus dont le génome est non-segmenté sont regroupés dans l'ordre des *Mononegavirales* sous la forme de quatre familles :

- *Paramyxoviridae* (virus de la rougeole, Sendai, Nipah, virus respiratoire syncytial)
- *Rhabdoviridae* (virus de la rage, de la stomatite vésiculaire)
- *Filoviridae* (virus Marburg, Ebola)
- *Bornaviridae* (virus Borna).

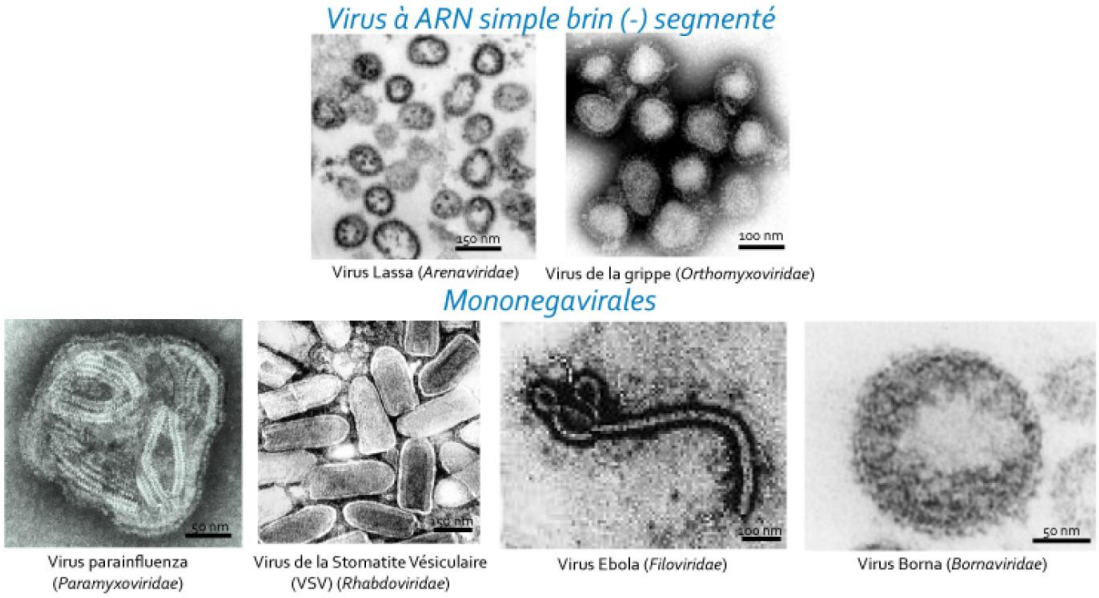


Figure 2: Images en microscopie électronique de particules virales de virus à ARN négatif. Les images proviennent de la galerie d'images de l'ICTVdB (International Committee on Taxonomy of Viruses database, (Buchen-Osmond 2003)). Virus Lassa (*Arenaviridae*), virus de la grippe (A/Hong Kong/1/68), virus de la stomatite vésiculaire (VSV), virus Ebola (Zaire Ebola, image prise en octobre 1976), virus Borna: courtoisie du Dr. F. A. Murphy (School of Veterinary Medicine, University of California, Davis. Virus parainfluenza: courtoisie du Dr. Linda Stannard (Department of Medical Microbiology, University of Cape Town, South Africa).

Les *Mononegavirales* regroupent un grand nombre de virus, certains connus depuis des siècles comme le virus de la rage, d'autres découverts plus récemment, les virus dits « émergents ». Ils s'adaptent rapidement à de nouveaux hôtes et sont responsables d'épidémies a forte mortalité, comme celle attribuée au virus Nipah en 1998 (Chua, Bellini et al. 2000; Chua, Lam et al. 2000).

1 – Les Rhabdoviridae

La famille des *Rhabdoviridae* comporte plus de 200 virus infectant des mammifères, des poissons, des crustacés, des insectes, des reptiles et des plantes. Parmi les cinq genres appartenant a cette famille, les genres *Vesiculovirus*, *Lyssavirus* et *Ephemerovirus* infectent des animaux alors que les genres *Cytorhabdovirus* et *Nucleorhabdovirus* infectent des végétaux. Il semblerait qu'un grand nombre d'entre eux soient véhiculés par des insectes

(Ammar and Hogenhout 2008; El-Desouky, Chi-Wei et al. 2009); (Hogenhout, Redinbaugh et al. 2003). Ainsi, différentes espèces de mouches sont impliquées dans la transmission inter-espèces de certains *Vesiculovirus*, des moustiques transmettent le virus de la fièvre éphémère bovine, et des pucerons sont responsables de la transmission de *Rhabdoviridae* aux végétaux. Deux sérotypes majeurs de *Vesiculovirus* existent, les sérotypes Indiana et New Jersey. Ces virus infectent généralement le bétail et provoquent des lésions vésiculaires bénignes sur les organes externes des animaux, entraînant parfois d'importantes conséquences d'ordre économiques. Il existe d'autres *Vesiculovirus* capables de provoquer des lésions similaires chez les animaux et d'infecter l'homme : le virus Piry (isolé chez un opossum au Brésil) (Brun, Bao et al. 1995; Wilks and House 2009), le virus Chandipura (initialement isolé chez l'homme en Inde) (Rao, Basu et al. 2004) et le virus Isfahan (isolé à partir de mouches et d'hommes en Iran) (Tesh, Saidi et al. 1977; Tesh, Saidi et al. 1977; Marriott 2005; Wilks and House 2009).

Le genre *Lyssavirus* est quant à lui divisé en sept génotypes (Tableau 2), sur la base d'analyses de séquence de la nucléoprotéine (Bourhy, Kissi et al. 1993; Badrane, Bahloul et al. 2001), mais aussi plus récemment sur la base d'analyses des génomes entiers (Delmas, Holmes et al. 2008). Le génotype 1 englobe toutes les souches de rage alors que les autres génotypes correspondent à des virus apparentés à la rage (*Lagos Bat* : génotype 2 ; *Mokola* : génotype 3 ; *Duvenhage* : génotype 4 ; *EBVL1* (*European Bat LyssaVirus 1) :* génotype 5 ; *EBLV2* : génotype 6 ; *ABLV* (*Australian Bat LyssaVirus) :* génotype 7). Tous les génotypes excepté le génotype 3, pour lequel l'espèce hôte reste encore inconnue (Sabeta, Markotter et al. 2007), ont des chiroptères comme réservoir naturel. Récemment, quatre génotypes additionnels, qui infectent des chauves-souris en Asie centrale et du sud-est, ont été ajoutés: virus Aravan, Khujand, Irkut et West Caucacian Bat (Kuzmin, Franka et al. 2008; Kuzmin, Niezgoda et al. 2008; Kuzmin, Wu et al. 2008). Un autre lyssavirus non classifié, le Shimoni Bat *Lyssavirus*, a été identifié au Kenya et caractérisé (Kuzmin, Mayer et al. 2010).

Phylogroup	Genotype	Species	Abbreviation (ICTV)[a]	Geographical origin	Potential vector(s)
Isolates characterized					
I	1	Rabies virus	RABV	Worldwide (except several islands)	Carnivores (worldwide); bats (Americas)
I	4	Duvenhage virus	DUVV	Southern Africa	Insectivorous bats
I	5	European bat lyssavirus type 1	EBLV-1	Europe	Insectivorous bats (*Eptesicus serotinus*)
I	6	European bat lyssavirus type 2	EBLV-2	Europe	Insectivorous bats (*Myotis* sp.)
I	7	Australian bat lyssavirus	ABLV	Australia	Frugivorous/insectivorous bats (*Megachiroptera*/*Microchiroptera*)
II	2	Lagos bat virus	LBV	Sub-Saharan Africa	Frugivorous bats (*Megachiroptera*)
II	3	Mokola virus	MOKV	Sub-Saharan Africa	Unknown
Isolates to be characterized as new genotypes					
–	–	Aravan virus	ARAV	Central Asia	Insectivorous bats (isolated from *Myotis blythi*)
–	–	Khujand virus	KHUV	Central Asia	Insectivorous bats (isolated from *Myotis mystacinus*)
–	–	Irkut virus	IRKV	East Siberia	Insectivorous bats (isolated from *Murina leucogaster*)
–	–	West Caucasian bat virus	WCBV	Caucasian region	Insectivorous bats (isolated from *Miniopterus schreibersi*)

[a] ICTV = International Committee on Taxonomy of Viruses.

Tableau 2: Classification des Lyssavirus. (d'après WHO Expert consultation on Rabies – WHO technical report series No. 931, Geneva, 2005).

2 – Le virus de la rage

La rage est une maladie très ancienne, décrite depuis l'Antiquité, dont l'origine infectieuse n'a été proposée qu'au début du XIX^{ième} siècle (Zinke 1804). Le terme « rage » dérive de la racine sanskrite *rabh* signifiant « agir violemment ». Le virus de la rage est le représentant canonique du genre *Lyssavirus*, et sert de modèle d'étude pour les autres membres de la famille des *Rhabdoviridae*.

La rage est une zoonose virale affectant les mammifères terrestres et qui est très répandue dans le monde (plus de 2,5 milliards de personnes vivent dans des régions endémiques de rage) (Haupt 1999) (Figure 3). Il existe deux cycles épidémiologiques majeurs du virus de la rage : la **rage terrestre** et la **rage sauvage**.

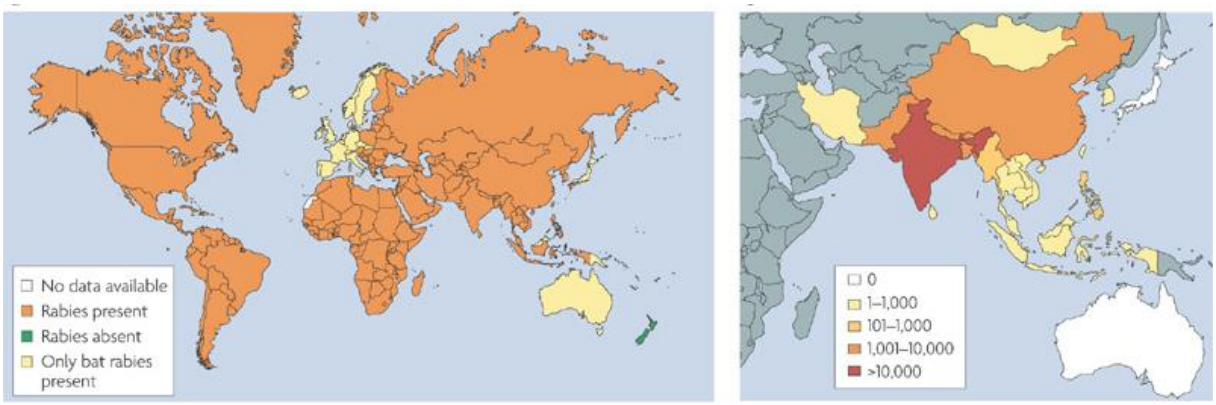


Figure 3: Répartition mondiale du virus de la rage (Schnell, McGettigan et al. 2010)**. A gauche**, la distribution mondiale du virus classique (génotype 1) en 2007 est représentée en orange. Les régions comprenant uniquement des Lyssavirus appartenant aux génotypes 2 à 7 sont colorées en jaune, et les régions exemptes de lyssavirus sont en vert. **A droite,** Le nombre de morts imputables à la rage en 2004 en Asie.

La rage terrestre (urbaine, des rues ou canine) est responsable de 90% des cas de rage dans le monde. Chez le renard, le chien et l'homme, la rage est un virus neurotrope qui provoque une infection du cerveau toujours mortelle lorsqu'elle est déclarée. La rage terrestre est transmise à l'homme principalement par les chiens qui constituent le vecteur principal du virus dans le monde. Chez l'homme, les contaminations rabiques sont le plus souvent dues à un contact (morsure) avec la salive de l'animal enragé, toutefois, le virus peut pénétrer l'organisme à la faveur de griffures, d'excoriations cutanées ou de léchage d'une plaie ouverte. En effet, le virus de la rage ne pénètre pas les téguments intacts, mais peut franchir les muqueuses saines. D'autres contaminations plus exceptionnelles ont été décrites, notamment liées à des greffes d'organes provenant de donneurs contaminés (Gode and Bhide 1988) (Srinivasan, Burton et al. 2005). Pendant la période d'incubation, le virus se propage dans l'organisme en cheminant le long des neurones **par voie axonale centripète** (Tirawatnpong, Hemachudha et al. 1989; Mrak and Young 1994). Cette période a une durée

variable, dépendante de l'inoculum et de la distance entre le site de morsure et le système nerveux central (SNC) du sujet. Le virus remonte le long des nerfs périphériques puis de la moelle épinière pour finalement aboutir au cerveau. Dans une deuxième phase, le virus se répand par **voie axonale centrifuge** vers divers organes préférentiels tels que les reins, la cornée et les glandes salivaires. Chez les personnes décédées par suite d'une infection par le virus de la rage, de hauts titres viraux sont retrouvés dans ces organes ainsi que dans les muqueuses. Le virus de la rage possède une forte affinité pour les zones du cerveau impliquées dans le comportement, ce qui induit souvent un comportement agressif chez l'animal ou l'homme, permettant ainsi la transmission du virus. L'animal contaminé, devenu enragé, va lui-même disséminer le virus à un autre hôte par morsure (Bleck and Rupprecht 2000).

Deux formes cliniques de la rage existent chez l'homme (Hemachudha, Laothamatas et al. 2002). Sur le plan physiopathologique, les deux formes de la maladie se distinguent par une inflammation modérée associée à une invasion neuronale sévère et rapide observée pour la forme furieuse, alors que la forme paralytique est caractérisée par une propagation plus lente au sein des neurones du patient, couplée à une réponse inflammatoire plus intense (Laothamatas, Wacharapluesadee et al. 2008). La **forme spastique dite rage furieuse** représente 70 % des cas. Les symptômes incluent troubles du comportement, hyperactivité, hallucinations et hydrophobie (impliquant aérophobie, dysphagie et salivation abondante). La phase aigüe dure de 2 à 7 jours puis le patient entre dans un coma et meurt par paralysie du système cardiorespiratoire en environ une semaine. La **forme paralytique dite rage muette** représente 30 % des cas. Seul 50% des personnes infectées présentent les symptômes d'hydrophobie et d'aérophobie, mais des paresthésies apparaissent au niveau de la région de la morsure, suivies de paralysie puis de paraplégie ou quadriplégie. La mort se produit par paralysie respiratoire au bout de 14 jours. La rage transmise à l'homme par les chiroptères peut donner lieu à des symptômes correspondant à la fois aux formes furieuses et paralytiques (Roine, Hillrom et al. 1988).

La rage sauvage (ou selvatique) des mammifères et des chiroptères touche de nombreuses espèces de chauve-souris (McColl, Tordo et al. 2000): par exemple, la grande chauve-souris marron (*Eptesicus fuscus),* la chauve-souris à tête d'argent (*Lasionycteris noctivagans*) et aussi la chauve-souris pipistrelle de l'est de l'Amérique (*Pipistrellus subflavus*). Certains chiroptères ainsi que d'autres mammifères sauvages (chacals, hyènes, mangoustes…), mais également des chiens Nigérians et Ethiopiens ayant faits l'objet de tests semblent être des porteurs asymptomatiques de la rage ainsi que d'autres lyssavirus (Warrell and Warrell 2004). Le virus chiroptérique est très rarement transmis aux humains, sauf lors de contact direct avec les animaux ou par contact avec des aérosols dans les grottes, où vivent les chauves-souris (Fooks, Brookes et al. 2003). Les espèces insectivores sont des vecteurs de rage dans toutes les régions du globe y compris dans les pays indemnes de rage terrestre tels que la Grande-Bretagne et l'Australie. En Amérique latine, ce sont surtout les chauves-souris hématophages (*Desmodus Rotundus*) (figure 4) qui transmettent la rage au bétail (Kobayashi, Sato et al. 2008). Les chiroptères, pouvant excréter le virus dans leur salive pendant de longues périodes sont des agents de contamination potentiellement redoutables et impossibles à vacciner.

Divers animaux sauvages constituent un réservoir permanent de la maladie. A leur contact, les animaux domestiques, les chiens errants voire l'homme peuvent être contaminés. Selon les régions du monde considérées, de nombreux vecteurs différents sont retrouvés, tels que le chacal et la mangouste en Afrique, la mangouste et le coyote en Amérique du Sud, le raton-laveur et la mouflette en Amérique du Nord, le renard au Moyen Orient, le loup en Europe occidentale, de l'Est et du Nord, le renard roux en Europe de l'Est et du Nord mais aussi le chien viverrin en Europe. Le virus est présent dans la salive des animaux infectés 3 a 5 jours avant l'apparition des symptômes neurologiques. L'animal meurt dans les 15 jours qui suivent la déclaration de la maladie. Il est probable que des *lyssavirus* infectant les chauve-souris soient entrés en contact avec les populations de renards roux eurasiens et aient établi un cycle réplicatif dans ce nouvel hôte par une série de mutations il y a plusieurs siècles, comme le suggèrent des analyses phylogénétiques réalisées sur la glycoprotéine de surface responsable de l'entrée du virus dans les cellules (Badrane and Tordo 2001). Une adaptation similaire pourraient avoir eu lieu aux Etats-Unis, où la rage terrestre affecte les ratons laveurs et les mouflettes (Rupprecht, Smith et al. 1995).

Figure 4 : La chauve-souris vampire *Desmodus Rotundus.* A gauche, en dessin et à droite, en photo.

Diagnostic de la rage

Adelchi Negri fut le premier à identifier en 1903 la présence d'ultrastructures visibles dans le cytoplasme des cellules, correspondant à l'accumulation de protéines virales et que l'on a appelées les corps de Negri (Miyamoto and Matsumoto 1965). Cependant, la détection de ces structures n'est pas fiable a 100% et d'autres techniques de détection permettant un diagnostic efficace et rapide telles que l'immunofluorescence directe, la culture cellulaire, les tests immunoenzymatiques ainsi que la détection par RT-PCR sont actuellement utilisées (Jackson 2009).

Prévention, vaccination et traitement de la rage

Il n'existe pas de traitement curatif de la rage. La vaccination représente donc la seule protection efficace contre le virus. Le vaccin original contre le virus de la rage fut développé par Louis Pasteur à partir de moelles épinières de lapin. Cependant, les vaccins préparés à partir de tissu cérébral, encore utilisés dans les pays en voie de développement, est faiblement immunogène et encéphalitogène du fait de la présence de myéline et d'impuretés dans les préparations (Bahri, Letaief et al. 1996; Shayam, Duggal et al. 2006). Aujourd'hui, des progrès considérables ont été réalisés dans la préparation des vaccins antirabiques. Ceux-ci sont mis au point en culture cellulaire (*human diploid cell vaccine* (HDCV), *rabies vaccine absorbed (rhesus monkey cells)* (RVA)) ou sur oeufs embryonnés de poussins (*purified ckick embryo cell vaccine* (PCEC)). Le vaccin est ensuite concentré puis inactivé. Le plus souvent l'injection est intramusculaire. La vaccination pré-exposition n'est utilisée que pour les personnes en contact avec le virus comme les vétérinaires, les fermiers et le personnel de laboratoire. Le traitement post-exposition consiste en une vaccination et une injection d'immunoglobuline antirabique (*rabies immunoglobulin* (RIG)) afin d'anticiper la réponse immunitaire de la personne infectée. Ce traitement reste cher mais efficace lorsque la maladie est traitée à temps. Cependant, il n'existe aucun médicament pour empêcher la multiplication virale et aucun traitement efficace une fois que les premiers symptômes sont apparus (Wilde 1999).

En Europe, la vaccination systématique (vaccination orale avec des vaccins atténués) des chiens puis des animaux sauvages (renards) a largement fait régresser la rage terrestre (Cliquet and Aubert 2004; Pastoret and Brochier 2009) Toutefois, en 2006, près de 8500 consultations pour suspicion de rage ont été réalisées en France et le nombre de traitements post-exposition chez des voyageurs mordus en zone d'endémie a plus que doublé en 20 ans (Institut Pasteur, http://www.pasteur.fr/ip/easysite/go/03b-000016-07j/rage-en-france2007). Quelques cas de guérisons inattendues ont été décrits (Alvarez, Fajardo et al. 1994; Madhusudana, Nagaraj et al. 2002; Is 2005), notamment le cas de Jeanna Giese, adolescente de 15 ans vivant dans le Wisconsin (Etats-Unis) qui a survécu à une infection par un variant de la rage associé à la chauve-souris à tête d'argent (*Lasionycteris noctivagans*), grâce a un nouveau traitement (Willoughby Jr, Tieves et al. 2005). Les médecins l'ont plongée dans le coma (*kétamine et midazolam*) et lui ont administré des antiviraux (*ribavirine et amantadine*) et des interférons afin de stimuler son immunité innée. Même si la jeune fille a gardé des troubles de la parole et de la stature plusieurs mois après son hospitalisation, elle a survécu. Malheureusement, de nombreuses tentatives infructueuses de traitements de patients atteints par la rage et d'autres lyssavirus transmis par les chauve-souris ont montrés depuis lors la non-reproductibilité du « Milwaukee Protocol » (Hemachudha, Sunsaneewitayakul et al. 2006; van Thiel, de Bie et al. 2009).

D'après les données de l'Organisation Mondiale de la Santé (OMS), la rage reste un problème important de santé publique puisqu'elle cause plus de 50 000 décès par an dans le monde. Même si toutes les classes d'âge sont sensibles, les enfants de 5 à 15 ans sont particulièrement exposés, puisque plus de 50 % des sujets mordus par des chiens enragés sont des enfants, le plus souvent des garçons de moins de 10 ans. L'incidence de la rage, humaine

et animale, est bien plus importante dans les pays en voie de développement. L'Asie concentre la majorité de ces cas (Inde, Pakistan, Bengladesh et Birmanie), mais des centaines de personnes meurent aussi de la rage en Afrique (Ouganda, Togo, Ghana, Egypte, Maroc). Chaque année, 6 à 7 millions de personnes reçoivent un traitement post exposition. Quelques cas sont répertoriés en Amérique du sud, et des dizaines de cas sporadiques autochtones ou de cas importés sont signalés en Amérique du nord ou en Europe.

Pour traiter la rage, Il semble essentiel de bloquer le virus avant qu'il ne puisse pénétrer dans le SNC ou de bloquer sa multiplication massive dans le cerveau. La connaissance des détails structuraux des processus de réplication de ces virus pourrait aider au développement de nouvelles cibles thérapeutiques et antivirales.

Le virus de la rage comme outil pour l'étude du système nerveux central

Le caractère neurotrope du virus rabique, ses capacités de transport rétrograde trans-synaptique, la relative lenteur de sa progression au sein du système nerveux central combinés aux possibilités d'études histologiques basées sur l'immunomarquage fluorescent (Ugolini 1995) en font un outil de choix pour diverses études neuroanatomiques, permettant notamment de mettre en évidence l'architecture des connections entre chaînes neuronales (Kelly and Strick 2000). Ces propriétés ont été utilisées avantageusement dans l'étude de certains processus neurologiques (Astic, Saucier et al. 1993; Tang, Rampin et al. 1999; Graf, Gerrits et al. 2002; Morcuende, Delgado-Garcia et al. 2002; Kelly and Strick 2004; Moschovakis, Gregoriou et al. 2004; Ruigrok, Pijpers et al. 2008; Ohara, Inoue et al. 2009).

3 – Le virus de la stomatite vésiculaire (VSV)

Cette maladie du bétail a été décrite aux Etats-Unis au XIX^{ième} siècle (Hanson 1952). Elle se localise principalement en Amérique du Nord et du Sud, toutefois, des épidémies de plus faible ampleur ont été décrites chez des chevaux en France et en Afrique du Sud à la fin du XIX^{ième} et au début du XX^{ième} siècle. Le virus a initialement été isolé par W. E. Cotton (Cotton 1927) et 1a morphologie a été établie par Chow et ses collaborateurs (Chow, Chow et al. 1954).

Le virus de la stomatite vésiculaire (VSV) touche principalement le bétail, les chevaux, et les porcs, et cause une maladie bénigne qui se caractérise par l'apparition de fièvre, de salivation excessive et de lésions vésiculaires sur les gencives, la langue, la bouche, sur les pis (vaches laitières) et les sabots de ces animaux. Cette maladie a des symptômes similaires à ceux de la fièvre aphteuse (*Picornaviridae, Aphtovirus*) et de la maladie vésiculaire du porc (*Picornaviridae, Enterovirus*). Les animaux éprouvent des difficultés à se nourrir et se mouvoir entraînant une perte de poids et un arrêt de la lactation chez les vaches laitières. Cette maladie a des conséquences économiques importantes dans les grandes fermes américaines. La maladie est rarement fatale sauf chez les porcs. Les épidémies de stomatite vésiculaire apparaissent soudainement (approximativement en 2 semaines) et se propagent rapidement dans le troupeau, mais pas nécessairement aux fermes adjacentes. Les tentatives

de développement de vaccins à base d'ADN ou en utilisant la glycoprotéine du virus comme agent immunogène ont donné des résultats médiocres (Yilma, Breeze et al. 1985; Cantlon, Gordy et al. 2000). Récemment, un vaccin constitué de virus atténué, obtenu par altération du gradient d'expression génique naturel du virus a été développé et testé sur des porcs avec des résultats satisfaisants (Flanagan, Zamparo et al. 2001).

L'infection est transmise par voie cutanée ou par contact avec des muqueuses infectées et le virus est retrouvé dans les fluides des vésicules et dans l'épithélium. Il ne persiste pas après la guérison. Le virus n'est excrété ni dans les urines, ni dans les fèces, ni dans le lait. La diffusion par contact direct avec les fluides est possible, et le transfert des lésions des mamelons par le matériel de tirage de lait est occasionnellement signalé. Cependant, les études expérimentales n'ont pas pu démontrer une transmission directe d'animal à animal, et on a remarqué de nombreuses épidémies où la maladie ne se propageait pas aux animaux voisins dans une exploitation (Sellers and Maarouf 1990). La maladie chez les animaux est brève, avec une virémie transitoire et minimale. La transmission par des arthropodes piqueurs du bétail semble être la plus probable. De plus, de nombreux *Vesiculovirus* ont été isolés chez les moustiques, les mouches, les moucherons, les mites et les tiques (Johnson, Tesh et al. 1969; Tesh, Chaniotis et al. 1972; Tesh, Boshell et al. 1987; Comer, Tesh et al. 1990). Il a été suggéré que certaines souches pouvaient être des souches de plantes, et que les animaux sont infectés directement, soit en mangeant les plantes infectées, soit en mangeant les insectes. Le virus peut parfois infecter les humains, causant généralement des symptômes de type grippaux et des lésions vésiculaires sur les lèvres et les mains (Johnson, Vogel et al. 1966), mais le plus souvent la maladie est asymptomatique. Au moins un cas humain d'encéphalite causée par le VSV a été documenté (Quiroz, Moreno et al. 1988). La période d'incubation est d'environ 9 jours après l'exposition et les symptômes peuvent persister 3 à 4 jours. De façon occasionnelle, il peut y avoir des rechutes. Toutefois, les infections ont été observées chez des personnes en contact avec un haut titre viral (inhalation d'aérosols, ou contamination accidentelle des yeux) comme les vétérinaires, le personnel travaillant dans les fermes en contact avec les animaux infectés ou encore les laborantins. Cependant, dans certaines zones de l'Amérique Centrale et du Nord où le VSV est endémique, il a été montré que la majorité des résidents ont développé des anticorps anti-VSV (Tesh, Boshell et al. 1987).

Aux Etats-Unis et en Amérique Centrale, ce sont principalement les sérotypes Indiana et New Jersey qui sont présents, alors qu'on trouve en Amérique du Sud, les souches Indiana Cocal et Alagoas. Les épidémies sont plus fréquentes dans les zones tropicales et subtropicales et apparaissent tous les ans ou tous les 2 à 3 ans, et les épidémies dans les zones tempérées apparaissent à des intervalles de 5 à 10 ans, pendant les mois d'été. Il existe une variation saisonnière pour la disparition de stomatite vésiculaire chez les animaux : elle disparaît à la fin de la saison des pluies dans les régions tropicales, et aux premières gelées dans les régions tempérées (Rodriguez 2002).

Le VSV comme outil thérapeutique

Le VSV constitue un excellent modèle pour la réplication et la transcription car c'est le virus le plus étudié de l'ordre des *Mononegavirales.* Il ne comporte que cinq gènes dont les signaux de régulation de l'expression sont plus simples que pour les autres virus de cet ordre.

Ce virus se multiplie facilement dans le cytoplasme de nombreuses souches cellulaires, et des tests de transcription et de réplication *in vitro* ont été établis. Il est aussi très largement utilisé comme outil de laboratoire pour étudier divers aspects de la biologie cellulaire, pour comprendre l'immunité innée et acquise. Depuis quelques années, son potentiel thérapeutique a été étudié: en effet, le VSV semble adapté comme agent de vaccination (Lichty, Power et al. 2004; Ramsburg, Rose et al. 2004; Bukreyev, Skiadopoulos et al. 2006; Palin, Chattopadhyay et al. 2007; Schwartz, Buonocore et al. 2007; Liao, Publicover et al. 2008; Roediger, Kugathasan et al. 2008) mais aussi comme virus oncolytique (Ebert, Shinozaki et al. 2003; Ebert, Harbaran et al. 2004; Lichty, Stojdl et al. 2004; Edge, Falls et al. 2008; Ozduman, Wollmann et al. 2008; Shi, Tang et al. 2009). Toutefois, peu d'informations sont disponibles quant à l'effet de l'administration de ce virus à l'homme, et les types de tissus infectés ne sont pas clairement établis.

II. La transcription et la réplication chez les Rhabdoviridae

A. Morphologie et composition des Rhabdoviridae

Les *Rhabdoviridae* sont des virus enveloppés dont la membrane lipidique dérive de la membrane de la cellule hôte infectée. Le génome de ces virus est composé d'une molécule **d'ARN simple brin de polarité négative** (complémentaire de l'ARNm). La particule virale des *Rhabdoviridae* possède une forme caractéristique évoquant une balle de revolver (« *bullet shape* ») (Figure 5 et 6). L'une des extrémités du virus est plate alors que l'autre est en forme d'ogive. Le diamètre des particules virales avoisine les 75 nm et leur longueur varie entre 100 et 300 nm (Howatson and Whitmore 1962). Ces virus sont composés à 74 % de protéines, 20 % de lipides, 3 % de carbohydrates et 3 % d'ARN (Thomas, Newcomb et al. 1985).

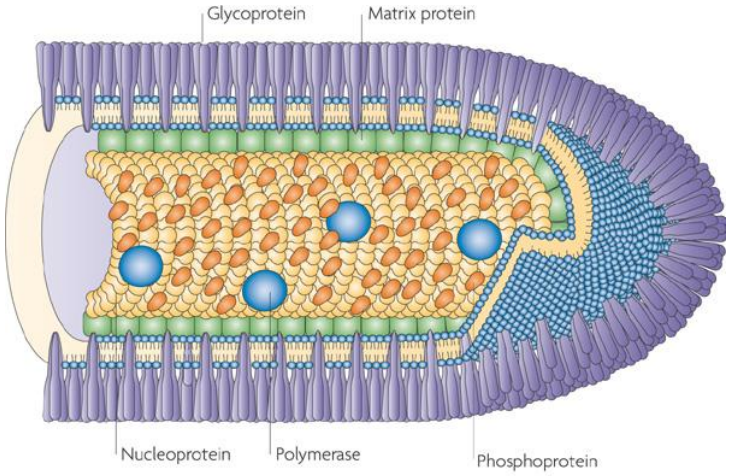


Figure 5: Représentation schématique de la particule virale en forme de balle de revolver des *Rhabdoviridae* (Schnell, McGettigan et al. 2010). Les protéines P et L, qui se fixent sur la face interne de la nucléocapside virale, sont néanmoins représentées sur la partie extérieure pour plus de clarté.

Les virus de l'ordre des *Mononegavirales* partagent la même organisation structurale et génomique, ainsi que des modes de réplication et de transcription similaires (Knipe 2006).

Il existe en tout **cinq protéines virales majeures.** La **glycoprotéine (G)** est enchâssée dans la membrane lipidique (David 1973) et forme des spicules trimeriques que l'on peut observer en microscopie électronique (Doms, Keller et al. 1987; Gaudin, Ruigrok et al. 1992). La **protéine matrice (M)** stabilise la forme condensée de la nucléocapside. La structure interne de la particule virale est constituée de trois protéines virales qui s'associent a l'ARN génomique afin de former un complexe hélicoïdal nommé **nucléocapside**, l'unité infectieuse du virus. L'ARN viral n'est jamais nu, ni dans les particules virales ni dans les cellules infectées, mais il est étroitement associé avec la **nucléoprotéine (N)** (Albertini, Wernimont et al. 2006; Green, Zhang et al. 2006). Le complexe nucléoprotéine-ARN sert de matrice pour la réplication et la transcription (Arnheiter, Davis et al. 1985). En effet, il a été montré que lors de l'infection, après l'entrée du virus dans la cellule, la synthèse des ARNm n'est possible qu'à partir de cette matrice nucléoprotéine-ARN viral. Deux protéines encodées par le génome viral sont nécessaires pour synthétiser l'ARN viral de façon efficace et régulée (Emerson and Yu 1975; Pattnaik and Wertz 1990): l'**ARN polymérase ARN dépendante virale (L)** qui possède les activités de transcriptase et réplicase, responsable de la synthèse des ARNs pendant le cycle viral, et son cofacteur, la **phosphoprotéine (P)**, essentielle pour la transcription et la réplication du génome viral.

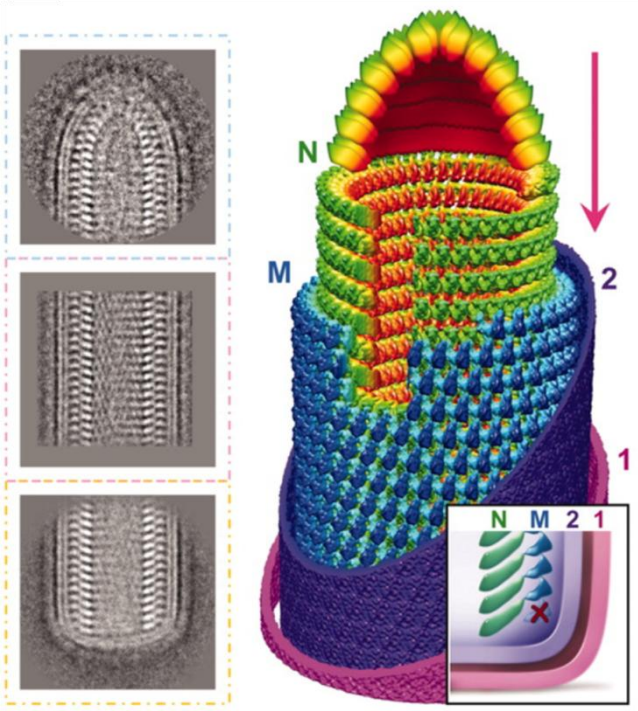


Figure 6: Architecture du virion de VSV (Ge, Tsao et al. 2010)**.** A gauche, moyennes de classes 2D représentatives de l'extrémité conique, du tronc et de la base du virion. A droite, modèle de l'extrémité conique et du tronc obtenu par reconstruction en cryo-microscopie électronique. La protéine N est représentée en vert, M en bleu, et les faces internes (2) et externes (1) de la membrane sont en violet et en rose, respectivement. En insert, illustration de la base du virion. Le "X" marque l'absence d'un tour d'hélice de M sous le dernier tour de l'hélice de N.

Récemment, la structure quasi-atomique des virions de VSV a été résolue par cryo-microscopie électronique (Ge, Tsao et al. 2010), révélant l'architecture fine du virion et l'agencement relatif des protéines N, M, et de la membrane lipidique au sein des particules virales (Figure 6). Malheureusement, les protéines G présentes à la surface du virion, ainsi que les protéines P et L associées aux nucléocapsides n'ont pas pu être reconstruites car elles ne respectent pas une symétrie hélicoïdale.

B. Génome des Rhabdoviridae

Le génome des *Rhabdoviridae* est constitué d'un unique brin d'ARN de polarité négative (Figure 7). Cet ARN viral code au minimum pour cinq protéines virales qui sont communes à tous les *Rhabdoviridae*, et dont les gènes sont organisés de façon similaire le long de la molécule d'ARN génomique (Figure 7). On trouve ainsi successivement les gènes codant pour la nucléoprotéine (N), la phosphoprotéine (P), la protéine matrice (M), la glycoprotéine (G) et l'ARN polymérase ARN-dépendante (L) (Tordo, Poch et al. 1986). La taille du génome est comprise entre 11 et 15 kb pour les *Rhabdoviridae*: 11 161 nucleotides (nt) pour le VSV, et 11 932 nt pour RV. Il existe des séquences spécifiques à l'extrémité 3' du génome (région *leader*) et à l'extrémité 5' (région *trailer)* (Blumberg, Leppert et al. 1981; Keene, Thornton et al. 1981; Blumberg, Giorgi et al. 1983).

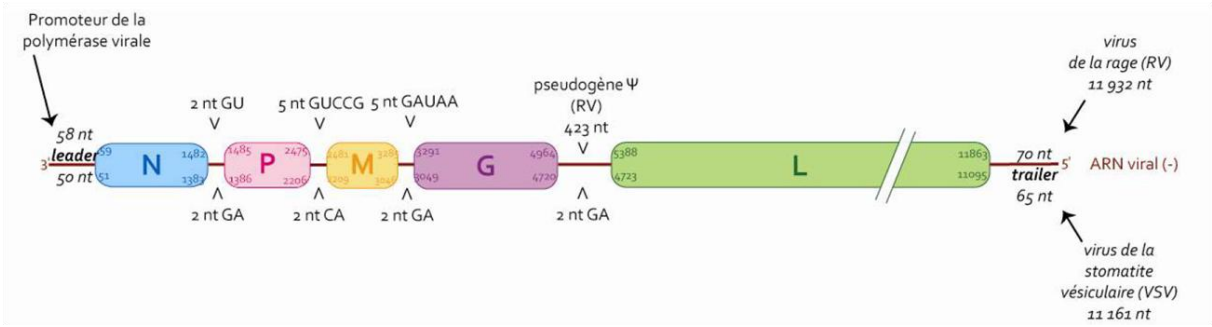


Figure 7 : Organisation schématique du génome des *Rhabdoviridae*, notamment du virus de la rage (RV) (*Lyssavirus*) et du virus de la stomatite vésiculaire (VSV) (*Vesiculovirus*). La longueur des séquences *leader* et *trailer* sont indiquées en nombre de nucléotides. La position de chaque gène est précisée. La localisation du pseudogène ψ est indiquée : cette séquence est variable suivant les souches de *Lyssavirus*.

Les gènes viraux sont séparés par des régions intergéniques de tailles variables pour RV, mais réduites à 2 nt pour VSV (Rose and Iverson 1979). Pour RV, les régions intergéniques entre les gènes N/P, P/M et M/G sont généralement courtes (2 à 5 nt) et conservées, par contre la région intergénique entre les gènes G et L est plus longue (423 nt) et de longueur variable chez les différentes souches de virus rabique. Cette séquence supplémentaire présente chez certains *Rhabdoviridae* dont le virus de la rage est nommée pseudogène ψ (Tordo, Poch et al. 1986). Chez les *Novirhabdovirus*, notamment chez le virus de la nécrose hématopoïétique infectieuse (IHNV, virus infectant les poissons), cette région constitue un sixième gène, codant pour une protéine NV (*Non-Virion*). Cette protéine de petite taille (12 kDa) est exprimée dans les cellules infectées (Kurath, Ahern et al. 1985; Kurath and Leong 1985) et constitue un facteur de virulence (Thoulouze, Bouguyon et al. 2004). Les mécanismes de l'expression du génome viral des *Mononegavirales* ont été largement explorés sur le modèle du virus de la stomatite vésiculaire mais aussi sur le virus Sendai (*Paramyxoviridae*) (Banerjee 1987; Tordo, Poch et al. 1988; Vidal and Kolakofsky 1989; Banerjee and Barik 1992; Tordo and Kouknetzoff 1993; Barr, Whelan et al. 2002; Kolakofsky, Le Mercier et al. 2004; Whelan, Barr et al. 2004; Grdzelishvili, Smallwood et al. 2005; Vulliemoz, Cordey et al. 2005; Li, Wang et al. 2006).

C. Le cycle viral

Les *Rhabdoviridae* ont un cycle viral exclusivement cytoplasmique, à l'exception des *Nucleorhabdovirus* dont le cycle viral présente une phase nucléaire (Martins, Johnson et al. 1998; Goodin, Austin et al. 2001; Goodin, Yelton et al. 2005; Deng, Bragg et al. 2007; Cisneros, Tsai et al. 2008; Bandyopadhyay, Kopperud et al. 2010). Certaines protéines virales pénètrent cependant dans le noyau pour interagir avec les partenaires cellulaires, et notamment inhiber la réponse immunitaire de la cellule hôte comme cela a été montré pour le virus de la rage (Blondel, Regad et al. 2002; Pasdeloup, Poisson et al. 2005; Chelbi-Alix, Vidy et al. 2006; Vidy, El Bougrini et al. 2007). Le cycle viral se décompose en trois étapes principales (Figure 8): l'entrée du virus (adhésion, endocytose, fusion membranaire et libération des nucléocapsides dans le cytoplasme), la multiplication du virus (transcription, traduction, réplication), et enfin l'assemblage et le bourgeonnement des nouveaux virions.

Entrée du virus

Le cycle viral débute par l'**adhésion** du virus à la membrane plasmique cellulaire médiée par les spicules de glycoprotéines présentes à la surface de la particule virale (Matlin, Reggio et al. 1982; Sun, Roth et al. 2010). Pour le virus de la rage comme pour le VSV, le tropisme cellulaire est large, permettant l'infection de nombreux types cellulaires, et donc de nombreux tissus (Reagan and Wunner 1985; Lichty, Power et al. 2004). Les récepteurs cellulaires impliqués dans l'internalisation du virus de la rage ne sont pas clairement définis, mais il semblerait que plusieurs récepteurs de type protéiques, lipidiques, ou glycosidiques soient impliqués (Superti, Seganti et al. 1984; Wunner, Reagan et al. 1984; Burrage, Tignor et al. 1985; Reagan and Wunner 1985; Conti, Superti et al. 1986; Seganti, Superti et al. 1986; Superti, Girmenta et al. 1986; Superti, Hauttecoeur et al. 1986; Lentz, Hawrot et al. 1987; Baer, Shaddock et al. 1990; Broughan and Wunner 1995; Gastka, Horvath et al. 1996; Thoulouze, Lafage et al. 1998; Tuffereau, Benejean et al. 1998; Tuffereau, Desmezieres et al. 2001). Pour VSV, il a été montré que l'endocytose était un processus clathrine- et dynamine-2 dépendant (Cureton, Massol et al. 2009), d'une durée inférieure à 3 minutes (Sun, Yau et al. 2005; Johannsdottir, Mancini et al. 2009), et probablement médié par l'interaction de la protéine G avec des lipides membranaires cellulaires (Carneiro, Bianconi et al. 2002; Carneiro, Lapido-Loureiro et al. 2006). Le processus de **fusion membranaire** entre les membranes cellulaires endosomales et virales est déclenché par l'acidification de l'endosome (White, Matlin et al. 1981; Gaudin 2000; Carneiro, Ferradosa et al. 2001; Roche and Gaudin 2004), libérant ainsi le complexe ribonucléoprotéique dans le cytoplasme de la cellule hôte. Chez VSV, les données les plus récentes montrent que la libération des RNPs s'accompagne de l'import nucléaire d'une fraction des protéines de matrice M, alors que la majeure partie des protéines M présentes au sein du virion reste liée à la membrane endosomale (Mire, Dube et al. 2009; Mire, White et al. 2010). Par ailleurs, la demi-vie du processus d'entrée et de libération des RNPs virales dans le cytoplasme a été estimée à 28 minutes, sur base d'analyses cinétiques *in vivo* de virions de VSV par imagerie de fluorescence (Das, Panda et al. 2009). Les données disponibles concernant le virus de la rage sont plus limitées, cependant la phase de latence de l'infection rabique est de l'ordre de quelques jours à plusieurs semaines, en

raison du lent processus de transport intracellulaire des virions jusqu'au système nerveux central (Klingen, Conzelmann et al. 2008).

Multiplication du virus

L'étape suivante est la **transcription**. L'ARN viral est transcrit en ARN messagers (ARNms) coiffés et polyadénylés par le complexe ARN polymérase ARN-dépendante–phosphoprotéine (Emerson and Yu 1975; Abraham and Banerjee 1976; Banerjee, Abraham et al. 1977).

Ces ARNms sont ensuite **traduits** en protéines virales par la machinerie cellulaire hôte. Les ARNms codant pour N, M, G et L sont traduits en protéines. Cependant, le gène codant pour la phosphoprotéine P chez le VSV peut être lu selon un cadre de lecture alternatif, permettant la traduction d'une protéine supplémentaire (Spiropoulou and Nichol 1993), un phénomène également bien documenté chez les *Paramyxoviridae* (Elliott, Yeo et al. 1990; Berg, Hjertner et al. 1992; Kurotani, Kiyotani et al. 1998). Chez le virus de la rage, l'initiation de la traduction est possible à différents loci sur l'ARNm encodant la protéine P, entraînant la production de plusieurs formes tronquées de la protéine (Chenik, Chebli et al. 1995).

Lorsque la quantité de protéines virales néo-synthétisées atteint un niveau suffisant, la polymérase virale entre en mode **réplicatif**. Chez VSV, il a été montré que les complexes protéiques contenant la protéine L et impliqués dans la transcription et la réplication étaient distincts (Qanungo, Shaji et al. 2004). Cependant, le mécanisme gouvernant le « switch » transcription/réplication n'est pas clairement défini (Perlman and Huang 1973; Perrault, Clinton et al. 1983; Barik and Banerjee 1992; Wertz, Whelan et al. 1994; Das, Pattnaik et al. 1997; Gupta and Banerjee 1997; Li and Pattnaik 1999; Whelan and Wertz 1999). En mode réplicatif, le complexe réplicase ignore les signaux de début et de fin de gène et synthétise une molécule d'ARN positif, l'antigénome, qui servira de matrice pour la synthèse de nouveaux génomes viraux. Ceux-ci peuvent alors servir soit de matrice pour la synthèse des ARNm viraux (transcription secondaire) soit pour la synthèse de nouveaux génomes viraux (réplication) (Banerjee 1987). Dès le début de leur synthèse, les génomes et antigénomes viraux sont encapsidés par la nucléoprotéine néo-synthétisée, qui joue un rôle structural critique (Patton, Davis et al. 1984; Moyer, Smallwood-Kentro et al. 1991).

Assemblage et bourgeonnement des particules virales

Pour l'**assemblage** des nouvelles particules virales, les ARN viraux néoformés associés à la nucléoprotéine, à la phosphoprotéine et à la polymérase virale sont exportés vers la membrane plasmique (David 1973; Hunt and Summers 1976). Les glycoprotéines sont transportées *via* l'appareil de Golgi (Bergmann, Tokuyasu et al. 1981; Strous, Willemsen et al. 1983; Lotti, Torrisi et al. 1992). Chez le virus de la rage, la protéine matrice se place sur la face interne de la membrane plasmique où elle interagit avec les queues cytoplasmiques des spicules de glycoprotéines (Mebatsion, Weiland et al. 1999). Pour le VSV, il a été montré que la queue cytoplasmique de la protéine G était nécessaire pour le bourgeonnement (Whitt, Chong et al. 1989), mais qu'il était possible de modifier extensivement sa séquence, suggérant l'absence d'interaction spécifique (Schnell, Buonocore et al. 1998). En effet, il semble que le

VSV constitue un système adapté pour l'expression d'antigènes de surface étrangers (Schnell, Buonocore et al. 1996; Kretzschmar, Buonocore et al. 1997; Geisbert, Jones et al. 2005), une propriété qui a été utilisée avantageusement dans la conception de nouveaux vaccins (Haglund, Leiner et al. 2002; Garbutt, Liebscher et al. 2004; Iyer, Pahar et al. 2009). Par ailleurs, la protéine M interagit également avec les nucléocapsides dont elle stabilise la condensation en hélice (Newcomb and Brown 1981; Mebatsion, Weiland et al. 1999; Ge, Tsao et al. 2010). De nouvelles particules virales compactes sont formées et **bourgeonnent** hors de la cellule hôte, à l'aide de protéines cellulaires (Harty, Paragas et al. 1999; Harty, Brown et al. 2001; Irie and Harty 2005).

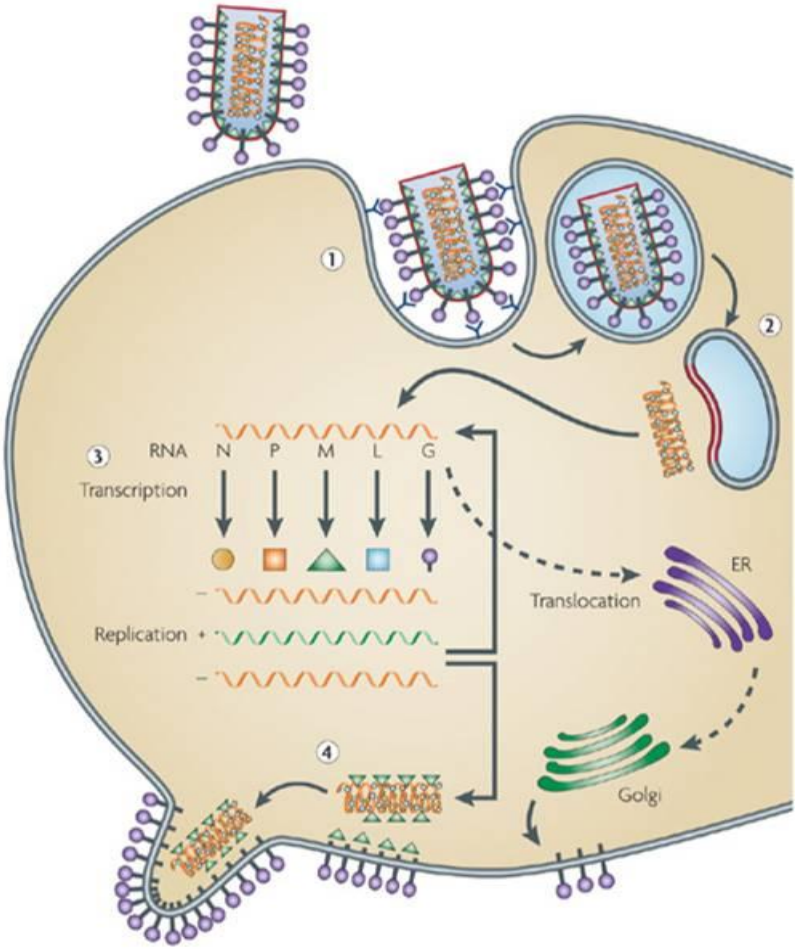


Figure 8 - Les trois phases du cycle viral des *Rhabdoviridae* (Schnell, McGettigan et al. 2010)**.** La première phase comporte la fixation au(x) récepteur(s) et l'internalisation du virion par la cellule hôte (1), suivie de la fusion des membranes virales et endosomales et de la libération dans le cytoplasme du matériel viral (2) (**étape 1**). Dans la seconde phase, les composants du virion sont produits (transcription, réplication et synthèse protéique (3) ; **étape 2**). La dernière phase du cycle viral consiste en l'assemblage des composants du virion et leur bourgeonnement hors de la cellule hôte (4) (**étape 3**). Les virions peuvent alors débuter un nouveau cycle infectieux.

1 – La transcription virale

Pour tous les virus à ARN négatif, la transcription constitue la première étape après l'entrée du virion dans la cellule hôte. Cette étape obligatoire précède la réplication virale et s'effectue de manière séquentielle (Abraham and Banerjee 1976; Ball and White 1976; Banerjee, Abraham et al. 1977). Sur la base des données obtenues chez le VSV, le déroulement de la transcription suit vraisemblablement le modèle dit du « *start-stop* » (Emerson 1982; Iverson and Rose 1982; Schnell, Buonocore et al. 1996; Stillman and Whitt 1998) (Figure 9). La transcription du génome viral est initiée au niveau d'un site unique d'entrée pour la polymérase virale situé a l'extrémité 3' du génome viral et l'accès au gène suivant dépend du signal de terminaison du précédent (Emerson 1982; Whelan and Wertz 2002). La transcription débute par la synthèse d'un ARN court de 48 nt pour le VSV et de 58 nt pour RV, appelé *ARN leader* (Colonno, Abraham et al. 1976; Colonno and Banerjee 1977; Colonno and Banerjee 1978; Tordo, Poch et al. 1986; Tordo and Kouknetzoff 1993), qui n'est ni coiffé ni polyadénylé et n'est donc pas traduit. Il a été montré que les nucléotides 19-29 et 34-46 situé dans cette région *leader* sont essentiels pour la transcription (Keene, Thornton et al. 1981; Whelan and Wertz 1999). Après la libération de l'*ARN leader*, la synthèse du premier ARNm, celui codant pour la protéine N, est initiée. La polymérase virale coiffe l'ARNm naissant dès l'initiation de la transcription. L'addition de la coiffe sur les ARNm viraux naissants s'effectue par un mécanisme propre aux *Rhabdoviridae*, identifié chez le VSV et qui a été extensivement étudié (Abraham, Rhodes et al. 1975; Moyer, Abraham et al. 1975; Testa and Banerjee 1977; Horikami and Moyer 1982; Barr, Whelan et al. 1997; Stillman and Whitt 1999; Gupta, Mathur et al. 2002; Grdzelishvili, Smallwood et al. 2005; Li, Fontaine-Rodriguez et al. 2005; Li, Wang et al. 2006; Li, Chorba et al. 2007; Ogino and Banerjee 2007; Li, Rahmeh et al. 2009; Rahmeh, Li et al. 2009; Ogino, Yadav et al. 2010). A la fin du gène codant pour la protéine N, la transcriptase atteint une séquence conservée : AUACUUUUUUU ($AUACU_7$), où elle commence à glisser de la matrice nucléotidique et produit une queue polyA à l'extrémité 3' de l'ARNm N. La polymérase va s'arrêter sur cette séquence consensus de terminaison AUAC et créer la polyA grâce à la répétition U_7 (Barr, Whelan et al. 1997). L'ARNm N mature, coiffé et polyadénylé est libéré. La transcriptase passe alors la région intergénique, et initie la transcription du deuxième ARNm, celui de P grâce à une séquence d'initiation conservée : UUGURRNGA (Iverson and Rose 1981). Ce processus est poursuivi pour les gènes suivants conduisant à la production de cinq ARNm coiffés et polyadénylés qui sont toujours produits dans un ordre séquentiel strict (Abraham and Banerjee 1976). La polymérase termine la synthèse environ 50 nt avant l'extrémité 5' du génome viral. L'atténuation qui se produit à chaque région intergénique crée ainsi un gradient d'ARNm produits (Iverson and Rose 1981; Finke, Cox et al. 2000). La transcription des gènes viraux est décroissante en fonction de leur position dans le génome.

En conséquence, la position relative des gènes permet de réguler la synthèse des protéines virales. Le gène de la nucléoprotéine est le plus proche du promoteur, assurant ainsi la synthèse de N en quantité suffisante pour l'encapsidation du génome viral. A l'inverse, le gène codant pour l'ARN polymérase ARN dépendante (L), nécessaire en quantités

catalytiques est situé en position distale (Abraham and Banerjee 1976). Des mesures cinétiques de l'activité transcriptionnelle chez le VSV ont montré que la polymérase réalise des pauses de 2,5 à 5,7 min au niveau de ces régions intergéniques (Iverson and Rose 1981). Ces pauses se déroulent lors des étapes de polyadénylation et d'addition de la coiffe sur les ARNm viraux, entraînant une transcription discontinue du génome. Dans 30% des cas, la transcription ne redémarre pas au niveau de ces régions intergéniques. Ce type de régulation est une caractéristique des *Mononegavirales*. La modification de l'ordre des gènes permet d'obtenir des virus viables, mais le rendement obtenu pour la multiplication de ces virus en culture cellulaire est beaucoup plus faible (Wertz, Perepelitsa et al. 1998; Finke, Cox et al. 2000).

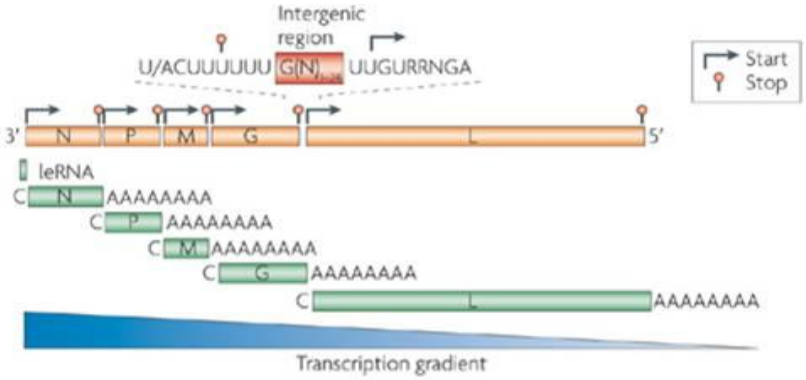


Figure 9: Représentation schématique de la transcription chez les *Rhabdoviridae* (Schnell, McGettigan et al. 2010)**.** Le génome de polarité négative encapsidé par la nucléoprotéine (orange) sert de matrice pour la transcription par la polymérase. La transcription commence par la synthèse d'un ARN leader non-coiffé (leRNA) à l'extrémité 3′ de l'ARN génomique, suivi par la transcription des ARNm coiffés en 5' (C) et polyadénylés en 3' (A) codant pour les protéines virales (vert). La polymérase s'arrête à la séquence signal (U/ACUUUUUUU), ignore la région intergénique de 2–24 nucléotides et reprend la transcription au niveau de la séquence signal (UUGURRNGA). Au cours de la transcription, l'*ARN leader* et les ARNm viraux codant pour les protéines virales s'accumulent suivant un gradient de concentration décroissant (bleu).

2 – La réplication virale

La réplication de l'ARN négatif nécessite la production d'une copie complémentaire du génome, appelée antigénome ou ARN complémentaire viral (ARNcv), qui est distinct des ARNm. L'ARN génomique, tout comme l'ARN antigénomique, est produit sous la forme d'un complexe ribonucléoprotéique. La polymérisation de l'ARN est associée à l'encapsidation. La réplication ne peut avoir lieu que lorsque la synthèse des nouvelles protéines virales est suffisante (Ball 2007). Au cours de la réplication, la séquence *leader* est encapsidée en premier (Blumberg, Giorgi et al. 1983; Tordo, Poch et al. 1986), et certaines données suggèrent que l'*ARN leader* contiendrait des signaux pour l'encapsidation (Yang, Hooper et al. 1998; Yang, Koprowski et al. 1999). Il a été montré que chez le VSV, l'*ARN leader* s'accumulait dans les noyaux des cellules infectées, inhibant la transcription de la cellule hôte (Kurilla, Piwnica-Worms et al. 1982). Cette propriété de l'*ARN leader* pourrait jouer un rôle dans le «switch » transcription/réplication du virus, dont le mécanisme précis

demeure à l'heure actuelle mal compris (Perlman and Huang 1973; Perrault, Clinton et al. 1983; Barik and Banerjee 1992; Wertz, Whelan et al. 1994; Das, Pattnaik et al. 1997; Gupta and Banerjee 1997; Li and Pattnaik 1999; Whelan and Wertz 1999). Cependant, les différentes études disponibles indiquent que le basculement entre les deux activités pourrait être régulé notamment par la quantité de nucléoprotéine déjà produite au cours du cycle virale (Arnheiter, Davis et al. 1985) et la quantité de complexe N-P présente (Gupta and Banerjee 1997), par des phosphorylations au niveau de la protéine P (Barik and Banerjee 1992; Barik and Banerjee 1992; Takacs, Barik et al. 1992), par l'étendue de la complémentarité des extrémités 5' et 3' de l'ARN génomique (Wertz, Whelan et al. 1994), ou encore par la présence de signaux spécifiques au niveau des régions *leader* (Li and Pattnaik 1999) et *trailer* (Whelan and Wertz 1999).

Dans le mode réplicatif, la polymérase virale ignore tous les sites d'initiation, de terminaison et de polyadénylation pour la synthèse des ARNm présents dans le génome viral. Les séquences présentes aux extrémités 3' et 5' de l'ARN viral génomique étant partiellement complémentaires (Wertz, Whelan et al. 1994), la polymérase virale peut amorcer la synthèse au niveau de plusieurs promoteurs :

- au niveau de la région *leader*, pour la transcription et la réplication,

- au niveau de *l'anti-trailer* de l'antigénome pour la réplication du brin positif. La séquence *trailer* de 69 nt et 64 nt respectivement chez RV et VSV, constitue à son tour un signal d'encapsidation par la nucléoprotéine (Bourhy, Tordo et al. 1989) lorsque la polymérase virale copie l'antigénome de façon processive.

La réplication se fait de manière asymétrique, et produit 10 fois plus de génomes que d'antigénomes (Whelan and Wertz 2002). En effet, la région *leader* est un promoteur moins fort que le promoteur situé sur la région *anti-trailer* (Whelan and Wertz 1999; Whelan and Wertz 1999). L'ARN antigénomique peut alors être répliqué par la polymérase virale pour produire plus de copies génomiques. Le nouvel ARN génomique peut être utilisé pour augmenter la transcription et la réplication de l'ARN viral ou plus tard dans l'infection pour produire de nouveaux virions.

D. Structures et fonctions des protéines virales

1 – Protéines associées à l'enveloppe virale

L'enveloppe virale est constituée à 50% en masse de lipides et à 50% de protéines (Pal, Barenholz et al. 1987; Stephens and Compans 1988). La composition lipidique de la membrane virale est relativement proche de la membrane cellulaire dont elle dérive (Kalvodova, Sampaio et al. 2009), mais elle montre cependant un enrichissement en

cholestérol (Luan, Yang et al. 1995). Cet enrichissement peut s'expliquer par un bourgeonnement du virus à partir de microdomaines lipidiques particuliers (Swinteck and Lyles 2008), tels que les radeaux lipidiques (*rafts*) (Pike 2009). Deux protéines virales sont impliquées dans la formation de l'enveloppe virale : la protéine de matrice (M), localisée à l'intérieur de la particule virale et la glycoprotéine (G), ancrée dans la membrane virale.

- *La glycoprotéine (G)*

La glycoprotéine est l'unique protéine membranaire intrinsèque des *Vesiculovirus* et *Lyssavirus*. Elle forme des spicules trimériques (Doms, Keller et al. 1987) visibles à la surface du virus par microscopie électronique (Gaudin, Ruigrok et al. 1992). La glycoprotéine joue de nombreuses fonctions au cours du cycle viral. Lors de l'entrée du virus dans la cellule hôte, elle assure les fonctions de reconnaissance du récepteur du virus, puis elle catalyse la fusion des membranes virales et cellulaires (Gaudin, Tuffereau et al. 1999; Gaudin 2000), permettant ainsi la libération du génome viral dans le cytoplasme de la cellule infectée. Chez le virus de la rage, il a été montré que la protéine G joue un rôle essentiel lors de l'assemblage des virions néo-synthétisés, en interagissant spécifiquement avec la protéine de matrice (M) (Mebatsion, Konig et al. 1996; Mebatsion, Weiland et al. 1999). La glycoprotéine porte des sites antigéniques responsables de la réponse immunitaire, capable d'induire la production d'anticorps (Benmansour, Leblois et al. 1991).

La nature des récepteurs impliqués dans la reconnaissance de la glycoprotéine G par la cellule hôte a fait l'objet de nombreuses études, notamment chez le virus de la rage, et constitue encore un sujet de débat à l'heure actuelle (Superti, Seganti et al. 1984; Wunner, Reagan et al. 1984; Burrage, Tignor et al. 1985; Reagan and Wunner 1985; Conti, Superti et al. 1986; Seganti, Superti et al. 1986; Superti, Girmenta et al. 1986; Superti, Hauttecoeur et al. 1986; Lentz, Hawrot et al. 1987; Baer, Shaddock et al. 1990; Broughan and Wunner 1995; Gastka, Horvath et al. 1996; Thoulouze, Lafage et al. 1998; Tuffereau, Benejean et al. 1998; Tuffereau, Desmezieres et al. 2001). Même si ces études ne définissent pas clairement la nature des récepteurs impliqués dans l'internalisation des particules virales, elles suggèrent que plusieurs récepteurs jouent probablement un rôle. Chez le VSV, la phosphatidylsérine a été proposée comme récepteur médiant la fixation du virus (Schlegel, Tralka et al. 1983). Bien que l'interaction de la protéine G du VSV à la surface des cellules semble relativement peu spécifique (Carneiro, Bianconi et al. 2002), des données récentes suggèrent que la présence de phosphatidylsérine augmente l'affinité pour la membrane cellulaire (Carneiro, Lapido-Loureiro et al. 2006).

La protéine G est synthétisée sous la forme d'un précurseur de 511 acides aminés (aa) chez VSV et 524 chez RV. Ce précurseur possède des aa en N-terminal qui constituent un peptide signal (16 pour VSV et 19 pour RV) pour l'insertion de la protéine dans le reticulum endoplasmique rugueux (RER) (figure 8) pendant la traduction. Ce peptide signal est clivé par la signalase dans le RER et la glycoprotéine acquiert sa longueur définitive de 495 aa chez

VSV et 505 aa chez RV. Par ailleurs, la protéine G subit de multiples glycosylations à l'intérieur de l'appareil de Golgi (McSharry and Wagner 1971; Schmidt and Schlesinger 1979; Rose, Adams et al. 1984). La protéine G mature est composée de trois domaines :

- La partie N-terminale de la protéine G constitue l'ectodomaine, exposé à la surface de la particule virale. Cet ectodomaine contient les sites de glycosylation et les déterminants antigéniques.

- Le domaine transmembranaire, majoritairement composé d'acides aminés hydrophobes, est replié en hélice α (aa 449 à 470 chez VSV, aa 440 à 460 chez RV).

- L'extrémité C-terminale de la protéine, qui constitue un segment cytoplasmique hydrophobe, peut interagir avec les autres protéines virales internes.

La glycoprotéine est responsable de la fusion de la membrane du virus avec celle de l'endosome. Le processus de fusion pH-dépendant a été étudié en détail chez le virus de la rage (Gaudin, Tuffereau et al. 1991; Gaudin, Ruigrok et al. 1993; Gaudin, Raux et al. 1996; Gaudin, Tuffereau et al. 1999; Gaudin 2000; Desmezieres, Maillard et al. 2003; Roche and Gaudin 2004) comme chez le VSV (Matlin, Reggio et al. 1982; Crimmins, Mehard et al. 1983; Carneiro, Ferradosa et al. 2001; Carneiro, Stauffer et al. 2003; Roche, Albertini et al. 2008; Sun, Roth et al. 2010).

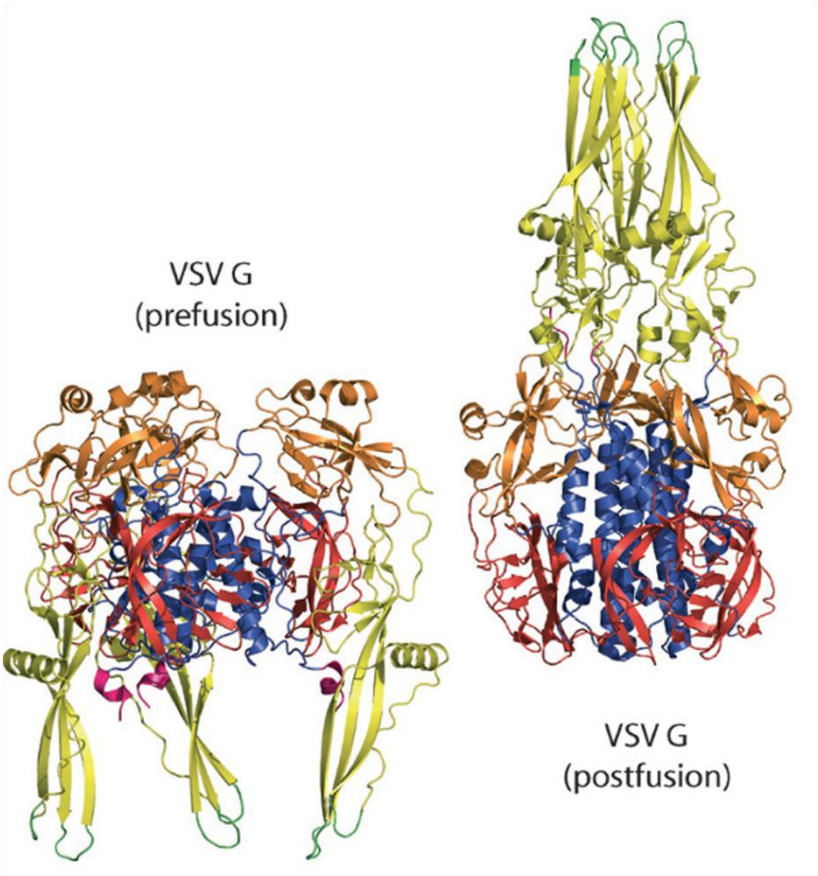


Figure 10: Structures trimériques de l'ectodomaine de la glycoprotéine du VSV dans ses conformations pré- et post-fusions (code pdb : 2j6j, résidus 1–413 et code pdb : 2cmz, résidus 1–410, respectivement). La glycoprotéine est coloré en fonction de ces domaines : domaine I (DI) : rouge ; domaine II (DII) : bleu ; domaine III (DIII) : orange et domaine IV (DIV) : jaune ; les boucles responsables de la fusion en vert, les zones visibles du domaine C-terminal en magenta (Roche, Albertini et al. 2008).

A la surface du virion, la protéine G est présente sous trois états conformationnels différents (Gaudin, Tuffereau et al. 1991; Gaudin, Ruigrok et al. 1993; Gaudin, Tuffereau et al. 1995; Roche and Gaudin 2002). L'état natif (N) est détecté à la surface des particules virales lorsque le pH est supérieur à 7. L'état actif (A) est présent immédiatement après l'acidification, et permet à la glycoprotéine de s'ancrer dans la membrane cible et d'amorcer la fusion (Durrer, Gaudin et al. 1995; Gaudin, Tuffereau et al. 1995). Après une incubation prolongée à pH acide (pH < 6,4), la protéine G est dans une conformation inactive pour la fusion (I) qui est antigéniquement distincte de l'état N (Gaudin, Tuffereau et al. 1995). Des résultats similaires ont été obtenus pour la protéine G du VSV (Clague, Schoch et al. 1990; Pak, Puri et al. 1997; Carneiro, Stauffer et al. 2003). Ces résultats indiquent que pour les *Rhabdoviridae*, la séquence des évènements aboutissant à la fusion membranaire est similaire à ce qui a été observé chez d'autres virus enveloppés (Hughson 1997; Weissenhorn, Dessen et al. 1999; Roche and Gaudin 2002; Roche, Albertini et al. 2008; White, Kielian et al. 2009).

Récemment, les structures cristallographiques de la glycoprotéine de VSV dans ses conformations pré- (Roche, Rey et al. 2007) (code pdb : 2j6j) et post-fusion (Roche, Bressanelli et al. 2006) (code pdb : 2cmz) ont été résolues, révélant d'importants changements conformationnels (Figure 10).

La structure atomique de la conformation post-fusion révèle que la protéine G du VSV est homologue à la glycoprotéine B du virus Herpes Simplex (*Herpesviridae*, virus à ADN) (Heldwein, Lou et al. 2006), et qu'elle combine des caractéristiques des protéines de fusions de classe I et de classe II. La structure post-fusion présente une conformation en feuillets β caractéristique des protéines de fusion de classe II, alors que le domaine de trimérisation possède une organisation en hélices α observée chez les protéines de fusion de classe I (Roche, Albertini et al. 2008).

- *La protéine matrice (M)*

Les protéines de matrice (M) des membres de la famille des *Rhabdoviridae* sont des protéines multifonctionnelles essentielles à l'assemblage et au bourgeonnement viral (Chong and Rose 1993; Finke and Conzelmann 2003). Elle stabilise la condensation hélicoïdale de la nucléocapside (Newcomb and Brown 1981). Elle est composée de 202 acides aminés (aa) pour le virus de la rage, 229 aa pour le VSV. Chez le VSV comme chez le virus de la rage, la protéine M subit diverses modifications post-traductionnelles qui semblent moduler sa fonction (Gaudin, Tuffereau et al. 1991; Kaptur, McKenzie et al. 1995; Harty, Brown et al. 2001). La protéine M est impliquée dans le contrôle du « switch » transcription/réplication (Finke, Mueller-Waldeck et al. 2003), et interfère avec les mécanismes de transcription et de traduction de la cellule hôte (Black and Lyles 1992; Ferran and Lucas-Lenard 1997; Her, Lund et al. 1997; Ahmed and Lyles 1998; Petersen, Her et al. 2000; von Kobbe, van Deursen et al. 2000; Glodowski, Petersen et al. 2002; Finke and Conzelmann 2003; Connor, McKenzie et al. 2006; Komarova, Real et al. 2007). L'extrémité N-terminale de la protéine M contient

plusieurs motifs « tardifs » riches en proline qui jouent un rôle essentiel dans le bourgeonnement viral (Harty, Paragas et al. 1999; Jayakar, Murti et al. 2000; Irie, Licata et al. 2004; Irie, Carnero et al. 2007; Wirblich, Tan et al. 2008). Des motifs similaires ont également été identifiés chez de nombreux autres virus (Dilley, Gregory et al. 2010; Liu, Cocka et al. 2010).

Pour le VSV, il a été proposé que la protéine M formerait des structures en forme de cigare autour desquelles la nucléocapside s'enrouleraient (Barge, Gaudin et al. 1993). La protéine M possède une forte tendance à l'oligomérisation, et forme des structures polymériques dans certaines conditions (McCreedy, McKinnon et al. 1990; Barge, Gagnon et al. 1996). Ces polymères ont une forme allongée caractéristiques, rappelant ces structures « en cigare » autour desquelles s'organise la nucléocapside virale (Barge, Gaudin et al. 1993). Il a été proposé que la polymérisation de la protéine M serait directement impliquée dans la compaction de la nucléocapside virale (Gaudin, Barge et al. 1995).

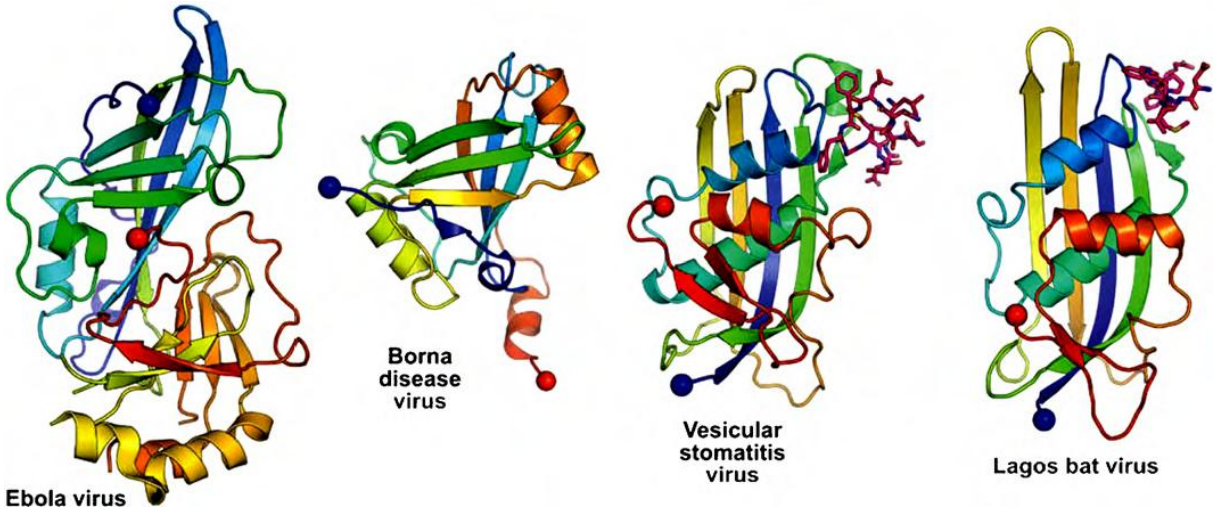


Figure 11: Structures atomiques des protéines matrices de différents virus à ARN négatif (Assenberg, Delmas et al. 2010)**.** Les structures des protéines matrice des virus Ebola (VP40, *Filoviridae*), Borna (*Bornaviridae*), VSV et Lagos Bat (*Rhabdoviridae*) sont représentées en cartoon et colorées du bleu (N-terminus) au rouge (C-terminus). Pour le VSV et Lagos Bat, les résidus localisés dans la région N-terminale et impliqués dans la polymérisation de la protéine sont indiqués en bâtonnets roses.

Pour le VSV et le virus Sendai (*Paramyxoviridae*), il a été montré que la protéine matrice interagissait en même temps avec la membrane et la nucléocapside (Chong and Rose 1993; Stricker, Mottet et al. 1994). La reconstruction tridimensionnelle de virions entiers obtenue récemment par cryo-microscopie électronique (Figure 6) fournit une représentation plus détaillée de cette double interaction (Ge, Tsao et al. 2010).

Les structures atomiques de plusieurs protéines matrices sont connues chez différents virus à ARN négatif (Figure 11): le VSV (Gaudier, Gaudin et al. 2002), le virus de la grippe (Sha and Luo 1997), le virus Ebola (Dessen, Volchkov et al. 2000), le virus Lagos Bat (Graham, Assenberg et al. 2008), et le virus de la maladie de Borna (Neumann, Lieber et al. 2009). La comparaison des structures cristallographiques des virus Lagos Bat et VSV révèle une organisation structurale de la protéine M qui semble conservée entre les *Lyssavirus* et les

Vesiculovirus, et ce malgré l'absence d'identité au niveau des séquences primaires (< 10%). Cependant, ces structures sont très différentes de celles des protéines matrice des virus appartenant aux familles *Filoviridae* et *Bornaviridae*, suggérant que ces protéines aux fonctions similaires sont apparues indépendamment au cours de l'évolution (Timmins, Ruigrok et al. 2004).

2 – Protéines impliquées dans la multiplication du virus et la synthèse d'ARN

- *L'ARN polymérase ARN-dépendante (L)*

Les premières ARN polymérases virales à avoir été caractérisées sont l'ARN polymérase ADN - dépendante du bactériophage T7 (Jones and Berg 1966; Summers and Siegel 1970; Chamberlin and Ring 1973), l'ARN polymérase ADN - dépendante du virus de la vaccine (Munyon, Paoletti et al. 1967) et l'ARN polymérase ARN-dépendante des virus de la famille des *Reoviridae* (Shatkin and Sipe 1968; Kalmakoff, Lewandowski et al. 1969; Lewandowski, Kalmakoff et al. 1969). L'activité ARN polymérase ARN - dépendante (L) du virus de la stomatite vésiculaire (VSV) fut mise en évidence peu de temps après à partir de virions purifiés (Baltimore, Huang et al. 1970). Chez les *Rhabdoviridae*, le nombre de molécules de protéine L par virion est faible, estimé à environ 50 molécules (Thomas, Newcomb et al. 1985), qui sont associées au complexe nucléoprotéine - ARN – Phosphoprotéine (Bishop and Roy 1972; Emerson and Wagner 1972; Emerson and Yu 1975). La polymérase est une protéine de 2142 aa chez le virus de la rage (Tordo, Poch et al. 1988) et de 2109 aa chez le VSV (Schubert, Harmison et al. 1984). Son gène représente ainsi 60% du génome viral (Figure 7). Cette protéine est la plus fortement conservée au sein des *Rhabdoviridae*, et même au sein de l'ordre des *Mononegavirales* (Tordo, Poch et al. 1988).

L'ARN polymérase ARN-dépendante (L) constitue la sous-unité principale du complexe enzymatique chez les *Mononegavirales*. Associée à son cofacteur, la phosphoprotéine (P), la polymérase assure à la fois les activités de transcription et réplication du génome viral, en utilisant exclusivement la matrice nucléoprotéine-ARN viral (Bishop and Roy 1972; Emerson and Wagner 1972; Emerson and Yu 1975). Elle est également responsable de la synthèse de la coiffe et de la polyadénylation des ARNm viraux (Banerjee, Moyer et al. 1974; Abraham, Rhodes et al. 1975; Moyer, Abraham et al. 1975; Testa and Banerjee 1977; Horikami and Moyer 1982; Barr, Whelan et al. 1997; Stillman and Whitt 1999; Gupta, Mathur et al. 2002; Grdzelishvili, Smallwood et al. 2005; Li, Fontaine-Rodriguez et al. 2005; Li, Wang et al. 2006; Li, Chorba et al. 2007; Ogino and Banerjee 2007; Li, Rahmeh et al. 2009; Rahmeh, Li et al. 2009; Ogino, Yadav et al. 2010).

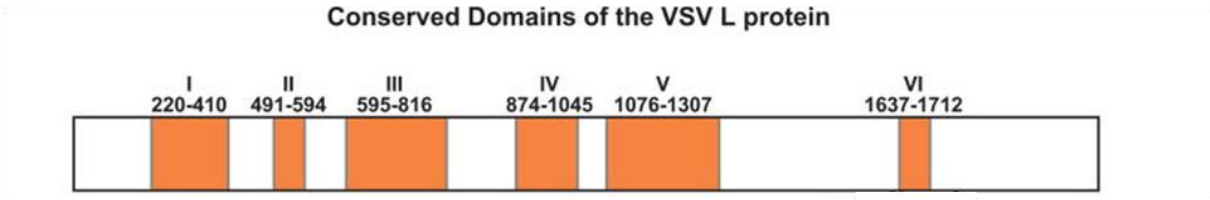


Figure 12 : Régions conservées de la polymérase (L) du VSV (Galloway, Richardson et al. 2008). Les domaines I-VI sont représentés en orange et leurs séquences sont indiquées d'après (Poch, Blumberg et al. 1990).

Six domaines particulièrement conservés ont été identifiés (Tordo, Poch et al. 1988; Poch, Sauvaget et al. 1989; Poch, Blumberg et al. 1990), ce qui a permis de repérer les sites probables d'interaction avec les matrices de réplication (Chenik, Schnell et al. 1998) et les nucléotides (Sleat and Banerjee 1993). Des analyses par mutagénèse tronquationnelle de la protéine L chez le virus de la rage et chez VSV ont en effet suggéré que le site d'interaction de l'ARN polymérase ARN-dépendante avec la phosphoprotéine (P) se situait dans la région C-terminale (Canter and Perrault 1996; Chenik, Schnell et al. 1998). Pour le virus Sendai, membre de la famille des *Paramyxoviridae*, la phosphoprotéine semble au contraire se fixer au niveau de la partie N-terminale (Holmes and Moyer 2002).

Peu d'informations structurales sont disponibles concernant les polymérases des virus à ARN négatif. Cependant, ces polymérases possèdent une certaine similarité de séquence avec les polymérases virales des virus à ARN positif, pour lesquels plusieurs structures sont connues, notamment celle du virus de l'hépatite C (Ago, Adachi et al. 1999; Lesburg, Cable et al. 1999) ou encore du virus de la Dengue (Yap, Xu et al. 2007) ou le virus West Nile (Yap, Xu et al. 2007)(*Flaviviridae*). Il semble que certaines de ces données structurales puissent s'appliquer aux polymérases des virus à ARN négatifs (Kolakofsky, Le Mercier et al. 2004). Par ailleurs, des informations structurales sont disponibles concernant la polymérase du virus de la grippe (*Orthomyxoviridae*, virus à ARN négatif segmenté) (Area, Martin-Benito et al. 2004; Torreira, Schoehn et al. 2007; He, Zhou et al. 2008; Dias, Bouvier et al. 2009). Récemment, un modèle structural par homologie a été obtenu *in silico* pour le domaine VI de la polymérase du VSV, responsable de l'activité 2'-O-ribose méthyltransférase, ce qui a permis une analyse à *posteriori* de certaines données de mutagénèse (Galloway, Richardson et al. 2008).

Additionnellement, des études réalisées sur le VSV suggèrent que les complexes transcriptases et réplicases, contenant tous deux la protéine L, sont distincts (Das, Pattnaik et al. 1997; Gupta, Shaji et al. 2003; Qanungo, Shaji et al. 2004), probablement en raison de l'association différentielle de multiples partenaires cellulaires dans ces deux complexes (Das, Mathur et al. 1998; Qanungo, Shaji et al. 2004).

- *La nucléoprotéine (N)*

La nucléoprotéine (N) est la gardienne du génome. Sa fonction première est d'encapsider le génome et l'antigénome viral, lui conférant ainsi une grande stabilité et une totale résistance aux ribonucléases (Simpson and Hauser 1966; Wagner, Schnaitman et al. 1969; Huang, Baltimore et al. 1970; Kiley and Wagner 1972; Soria, Little et al. 1974; Condra and Lazzarini 1980; Wagner and Rose 1996).

Dans les cellules infectées, on retrouve la nucléoprotéine sous deux formes :

- Sous sa forme monomérique, également appelée forme soluble, la nucléoprotéine forme un complexe de stœchiométrie 1 : 2 avec la phosphoprotéine. Ce complexe est appelé N^0-P, le signe « ^{0} » indiquant que la nucléoprotéine est libre d'ARN.
- Sous sa forme oligomérique, liée à l'ARN et formant la matrice ribonucléoprotéique utilisée pour la transcription et la réplication virale, et appelée nucléocapside ou complexe N-ARN.

La nucléoprotéine du virus de la rage est par ailleurs phosphorylée au niveau de la sérine 389 par la caséine kinase II (Wu, Lei et al. 2003), une kinase cellulaire ubiquitaire, lorsqu'elle est présente sous la forme de complexes N-ARN (Kawai, Toriumi et al. 1999). Ce phénomène de régulation par phosphorylation de la protéine N, qui module l'affinité du complexe N-ARN pour la phosphoprotéine (Toriumi, Honda et al. 2002; Toriumi and Kawai 2004), ne semble pas exister chez VSV.

- *Le complexe nucléoprotéine-phosphoprotéine N^0-P*

Le complexe N^0-P se forme transitoirement au cours du cycle viral, et joue un rôle critique lors de la transcription et de la réplication virale. La formation du complexe permet notamment d'empêcher la fixation non-spécifique de la protéine N à l'ARN lors de la transcription, alors que sa dissociation est nécessaire lors de la réplication afin d'encapsider les génomes et antigénomes néo-synthétisés. Lorsque l'on co-exprime dans un système de production recombinante la nucléoprotéine et la phosphoprotéine, une fraction de la nucléoprotéine forme un complexe N^0-P, où la nucléoprotéine est présente sous forme soluble et monomérique (Mavrakis, Iseni et al. 2003). L'ultracentrifugation analytique, la spectrométrie de masse et la microscopie électronique ont montré que ce complexe, pour le virus de la rage, était composé d'une nucléoprotéine associée à deux phosphoprotéines (Figure 13) (Mavrakis, Iseni et al. 2003). La nucléoprotéine de ce complexe soluble n'est pas phosphorylée au niveau de la sérine 389 (Kawai, Toriumi et al. 1999). En effet, la nucléoprotéine néo-synthétisée semble ne pas être immédiatement phosphorylée, mais elle

s'associe à la phosphoprotéine. Ce complexe N^0-P participe ensuite à la formation de complexes N-ARN.

Il semblerait que cette nucléoprotéine non phosphorylée fixe plus efficacement la matrice ARN (Yang, Hooper et al. 1998; Yang, Koprowski et al. 1999), et que sa phosphorylation ait lieu pendant ou après l'encapsidation (Toriumi, Honda et al. 2002). Pour le VSV, le complexe N^0-P soluble recombinant inhibe la transcription *in vitro* (Gupta and Banerjee 1997). Cependant, l'addition de complexe N^0-P provenant de cellules infectées à un système de transcription *in vitro* résulte en la synthèse du génome, laissant supposer que le complexe joue un rôle dans la régulation du basculement entre les activités de transcription et de réplication (Gupta and Banerjee 1997).

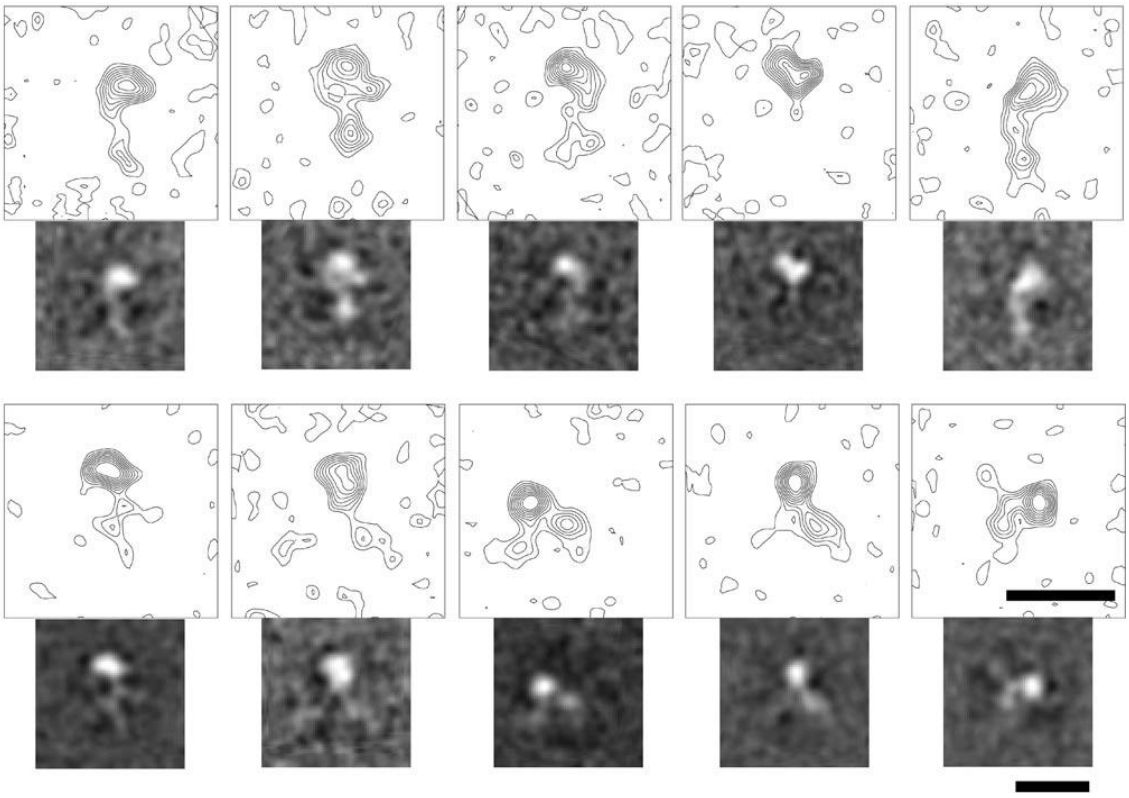


Figure 13: Images de microscopie électronique du complexe N_{His}^0-P_2 en coloration négative. Dix moyennes de classes d'images sont représentées (correspondant chacune à la moyennation d'environ 20 images brutes). Une représentation en lignes de niveau est dessinée au dessus de chaque classe d'images montrant plus clairement l'exclusion du colorant par le complexe. Dans la partie supérieure, les images montrent une unique extension de la densité partant de la nucléoprotéine, alors que les images de la partie inférieure affichent une double extension de la densité radiant de la nucléoprotéine. La barre d'échelle représente 10nm (Mavrakis, Iseni et al. 2003).

La nucléocapside

Au cours de la réplication virale, la nucléoprotéine s'associe à l'ARN viral pour former la nucléocapside, protégeant ainsi l'ARN viral des ribonucléases (RNases) tout en empêchant la formation de structures secondaires et l'hybridation des génomes et antigénomes viraux. Les nucléocapsides des virus à ARN négatifs possèdent une symétrie hélicoïdale (Compans and Choppin 1967; Sokol, Schlumberger et al. 1969; Egelman, Wu et al. 1989; Bhella, Ralph et al. 2002; Bhella, Ralph et al. 2004). Néanmoins, les paramètres hélicoïdaux, ainsi que la régularité des structures formées varient fortement en fonction des conditions expérimentales et du virus considéré, comme le montrent les clichés de microscopie électronique (figure 14).

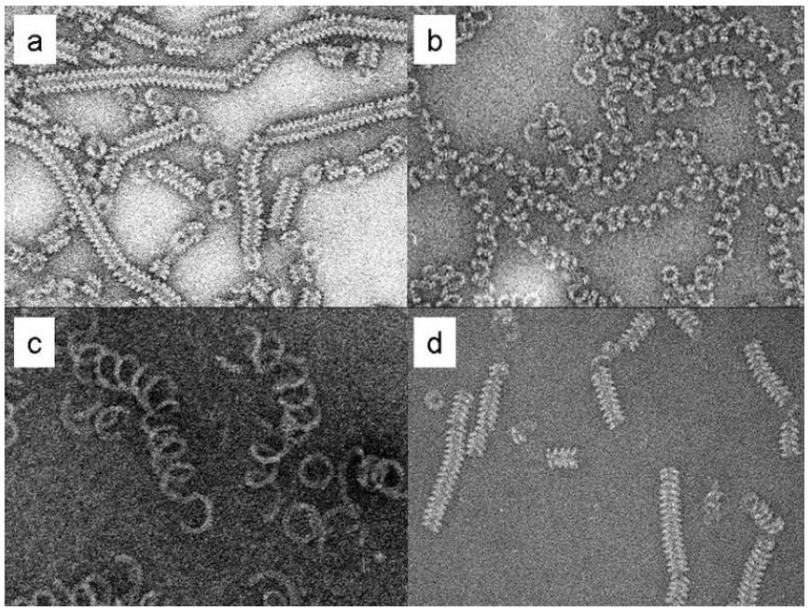


Figure 14: Images de microscopie électronique en coloration négative de quelques nucléocapsides de *Mononegavirales* (Albertini, Schoehn et al. 2005)**.** (a) Nucléocapsides recombinantes du virus de la rougeole (*Paramyxoviridae*). (b) Nucléocapsides virales du virus de la rage (*Rhabdoviridae*). (c) Nucléocapsides recombinantes du virus de Marburg (*Filoviridae*). (d) Nucléocapsides recombinantes du virus Nipah (*Henipahviridae*).

L'interaction des nucléoprotéines avec l'ARN est forte : il a notamment été montré que les nucléocapsides de P*aramyxoviridae* et de *Rhabdoviridae* sont très stables, et peuvent résister à de hautes concentrations salines et à une force de gravité importante (Simpson and Hauser 1966; Soria, Little et al. 1974; Condra and Lazzarini 1980; Heggeness, Scheid et al. 1980; Blumberg, Giorgi et al. 1984). Ainsi, les nucléocapsides résistent aux centrifugations en gradient de chlorure de césium (CsCl) en conditions hautement dénaturantes (8M d'urée)

(Iseni, Barge et al. 1998). Les changements importants de ces nucléocapsides en réponse à la force ionique indiquent la plasticité de cette structure hélicoïdale. Ces changements pourraient être importants lors de la transcription et de la réplication virale, ou lors du bourgeonnement, un processus dans lequel l'enroulement et la flexibilité de ces nucléocapsides changeraient pour adopter la structure hautement organisée présente à l'intérieur du virion (Figure 6) (Ge, Tsao et al. 2010). L'expression de la nucléoprotéine des *Mononegavirales* dans différents systèmes (eucaryotes ou procaryotes) produit des complexes dont la morphologie et les propriétés biochimiques sont identiques à celles des nucléocapsides virales (Iseni, Barge et al. 1998).

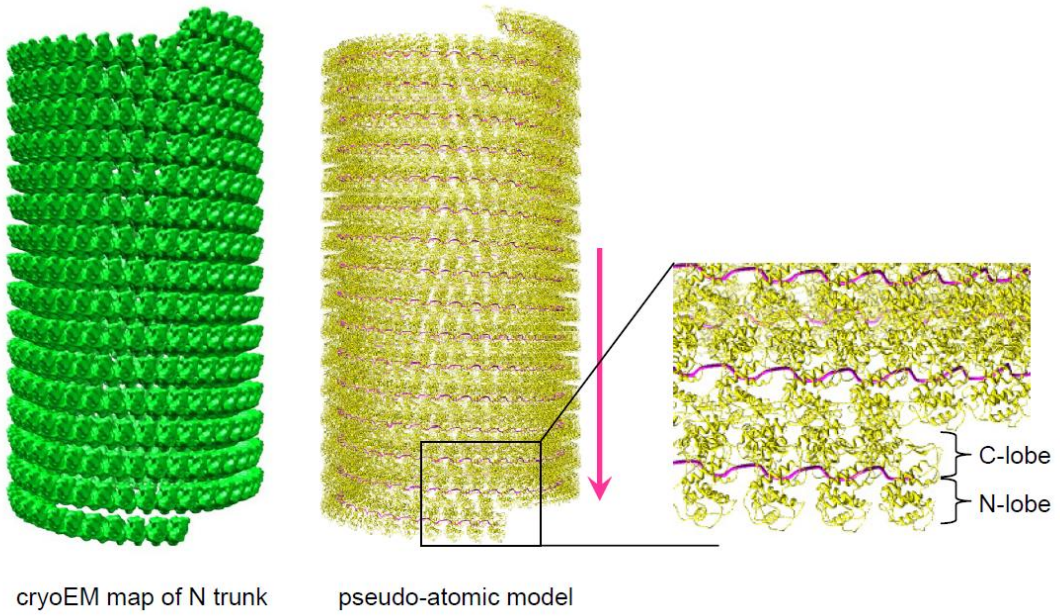


Figure 15: Organisation structurale de la nucléocapside au sein des virions du Virus de la Stomatite Vésiculaire (VSV) (Ge, Tsao et al. 2010)**.** A gauche, la densité électronique de cryo-microscopie correspondant à la nucléocapside extraite du tronc de la particule virale est représentée en vert (seuil 1.0 σ). Au centre, et en vue magnifiée à droite, le modèle pseudo-atomique généré par docking flexible du monomère cristallographique (code pdb : 2GIC) dans la densité expérimentale. La nucléoprotéine est représentée en cartoons jaunes et l'ARN est en magenta.

Des reconstructions hélicoïdales effectuées à partir d'images de microscopie électronique ont permis de déterminer les paramètres hélicoïdaux des nucléocapsides de certains virus, comme le nombre de sous-unités de nucléoprotéine par tour d'hélice. Pour le virus de la rougeole et de Sendai (*Paramyxovirus*), le nombre de nucléoprotéines par tour est de 13 (Egelman, Wu et al. 1989; Schoehn, Mavrakis et al. 2004). Pour le virus respiratoire syncytial humain (RSV) et le virus Simien 5 (SV5), ce nombre varie de 10 à 14 (Bhella, Ralph et al. 2002), soulignant une fois encore la plasticité de ces structures. L'hétérogénéité et la flexibilité de ces nucléocapsides excluent totalement leur cristallisation. De plus, la détermination des paramètres hélicoïdaux à partir de clichés de microscopie électronique nécessite une homogénéité et une rigidité qui n'ont pas été observées chez le virus de la rage (Figure 14b). Récemment, la reconstruction tridimensionnelle des virions de VSV en cryo-microscopie électronique a permis la visualisation à résolution quasi-atomique de la structure

des nucléocapsides au sein de la particule virale infectieuse (Ge, Tsao et al. 2010) (Figure 15), révélant notamment la présence de 37,5 sous-unités de nucléoprotéine par tour d'hélice au sein du tronc de la particule virale.

Fixation à l'ARN

La nucléoprotéine des virus à ARN négatif se fixe sur un nombre entier de nucléotides. Des études spectrophotométriques et de microscopie électronique ont permis de déterminer le nombre de nucléotides fixés par nucléoprotéine. Ce nombre est spécifique à chaque famille de virus au sein de l'ordre des *Mononegavirales*: la nucléoprotéine des *Paramyxoviridae* fixe six nucléotides (Egelman, Wu et al. 1989), celle des *Filoviridae* fixe entre 12 et 15 nt (Mavrakis, Kolesnikova et al. 2002) et celle des *Rhabdoviridae* en fixe neuf (Thomas, Newcomb et al. 1985).

Pour les *Paramyxoviridae*, à l'exception des *Pneumovirus* (Samal and Collins 1996), il existe une règle appelée **« règle de six »** (Calain and Roux 1993; Pelet, Delenda et al. 1996; Kolakofsky, Pelet et al. 1998; Phillips, Samson et al. 1998; Peeters, Gruijthuijsen et al. 2000; Halpin, Bankamp et al. 2004; Kolakofsky, Roux et al. 2005), qui indique que le nombre exact de nucléotides du génome viral doit être un multiple de six pour que la polymérase virale réplique efficacement le génome. Ceci signifie que chaque nucléoprotéine interagit avec six nucléotides, et que les sites d'interactions de la nucléoprotéine avec chacun de ces six nucléotides ne sont pas équivalents (Vulliemoz and Roux 2001; Iseni, Baudin et al. 2002). En revanche, aucune « règle des neuf » n'a été mise en évidence chez les *Rhabdoviridae* (Iseni, Baudin et al. 2000).

Les anneaux de nucléoprotéine

Lorsque les nucléoprotéines du virus de la rage ou du VSV sont exprimées en cellules d'insectes ou dans des bactéries, celles-ci encapsident des ARNs cellulaires de façon non-spécifique et forment de longues structures hélicoïdales de type nucléocapside (vraisemblablement en fixant des ARN messagers de la cellule), ainsi que des structures circulaires comportant un nombre variable de nucléoprotéines (9 a 15) (Figure 16) (Albertini, Clapier et al. 2006).

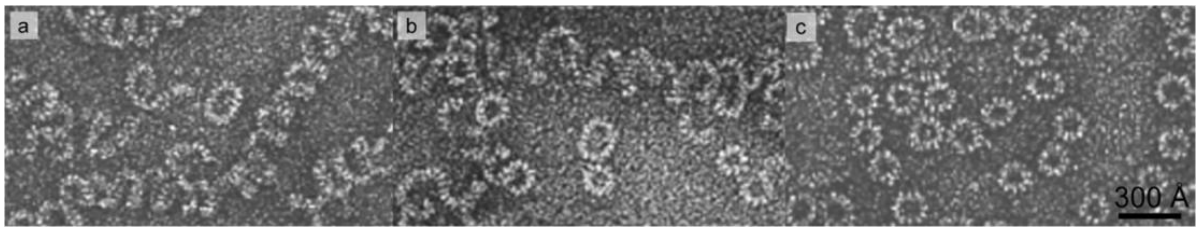


Figure 16: Les différents complexes nucléoprotéine-ARN du virus de la rage observés en microscopie électronique (coloration négative) (Albertini, Schoehn et al. 2008). (a) Nucléocapsides virales. (b) Nucléocapsides recombinantes produites en cellules d'insectes. (c) Anneaux issus de l'association de 9 à 15 sous-unités de nucléoprotéine et une molécule d'ARN simple brin. D'après (Iseni, Barge et al. 1998; Schoehn, Iseni et al. 2001; Mavrakis, Iseni et al. 2003; Albertini, Clapier et al. 2006; Albertini, Schoehn et al. 2008).

Différents travaux portant sur la nucléoprotéine du virus de la rage ont permis, d'une part, l'obtention d'une reconstruction tridimensionnelle des anneaux de nucléoprotéines à 10 sous-unités (N_{10}) à partir d'images de cryo-microscopie électronique (Figure 17), et d'autre part, la localisation du site de fixation de la phosphoprotéine sur base d'expériences de protéolyse limitée (Schoehn, Iseni et al. 2001). En effet, le traitement de ces anneaux par la trypsine entraîne un clivage au niveau du résidu 376 de la nucléoprotéine, éliminant ainsi une partie du domaine C-terminal, ce qui a pour effet d'inhiber la fixation de la phosphoprotéine (Schoehn, Iseni et al. 2001). La comparaison de la structure des anneaux N_{10} natifs et des anneaux protéolysés (Figure 17, i-l), qui ont perdus le domaine d'attachement de la phosphoprotéine, indique que la protéine P se fixe sur la partie supérieure de l'anneau.

Dans le cas du VSV, les travaux réalisés par Chen (Chen, Green et al. 2004) ont également permis d'obtenir une reconstruction à partir d'images de microscopie électronique des anneaux N_{10} du VSV, qui est relativement proche de celle des anneaux du virus de la rage.

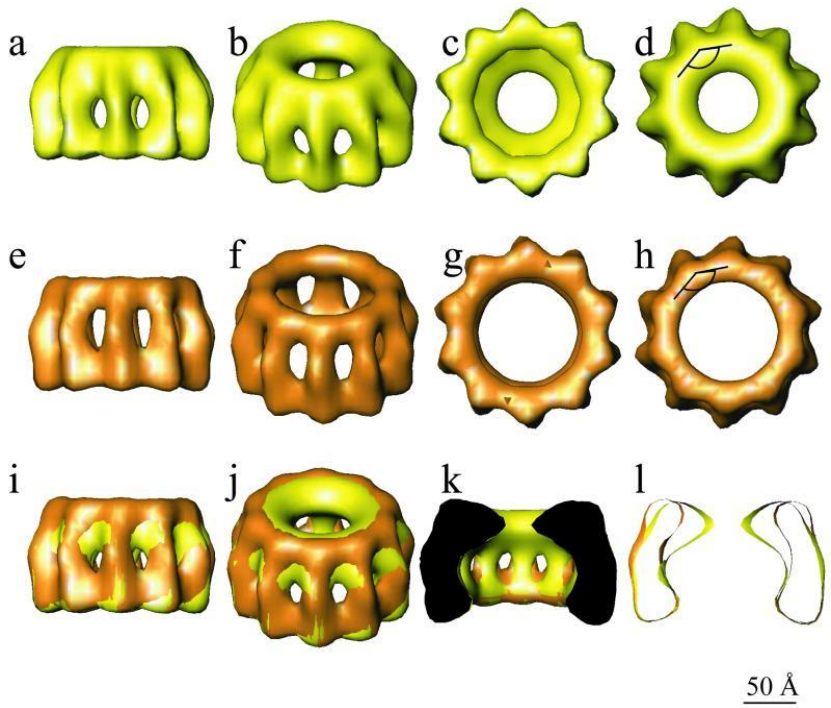


Figure 17: Reconstructions tridimensionnelles de complexes nucléoprotéine-ARN (N_{10}) du virus de la rage à partir d'images de cryomicroscopie électronique (Schoehn, Iseni et al. 2001)**.** (a-d) Différentes vues des anneaux N_{10} natifs sont représentées en jaune. (e-h) Les anneaux N_{10} digérés par la trypsine sont représentés en orange. (i-l) Superpositions des anneaux N_{10} natifs et digérés. Les deux structures sont très similaires. Une portion significative de la densité est absente au niveau d'une des faces pour les anneaux digérés (l). Dans les deux cas, l'angle mesuré entre les sous-unités est de 144°.

En 2006, la structure cristallographique des anneaux N_{11}-ARN du virus de la rage a été résolue à 3,5 Å (Albertini, Wernimont et al. 2006), et la structure cristallographique des anneaux N_{10}-ARN du virus de la stomatite vésiculaire (VSV) à 2,9 Å (Green, Zhang et al. 2006). Ces structures de nucléoprotéines sont les premières à révéler les détails atomiques de l'association avec l'ARN chez des virus à ARN négatif. Ces complexes ont tous deux été cristallisés dans le groupe d'espace $P2_12_12$, et les structures ont été résolues par SAD (*Single Anomalous Dispersion*). Pour le virus de la rage, les anneaux N_{11} (environ 550 kDa) sont positionnés face à face dans l'unité asymétrique et chaque anneau possède un diamètre externe de 160 Å, un diamètre interne de 60 Å et une hauteur de 74 Å (Figure 18, A & B).

Le protomère de nucléoprotéine consiste en deux domaines N et C-terminaux (NTD et CTD) qui sont en contact avec neuf nucléotides d'ARN simple brin (Figure 18 D & Figure 19 B). Le domaine N-terminal (NTD : résidus 32-233) est replié en 6 hélices α reliées par des longues boucles. Le domaine C-terminal (CTD : résidus 236-356 et 396-450) est composé de 11 hélices α liées par des boucles plus courtes. Deux régions sont flexibles, et absentes dans la structure cristallographique (NTD : 105 a 118 et CTD : 376 a 397). Les deux domaines NTD et CTD agissent comme des mâchoires en séquestrant complètement la molécule d'ARN (Figure 18).

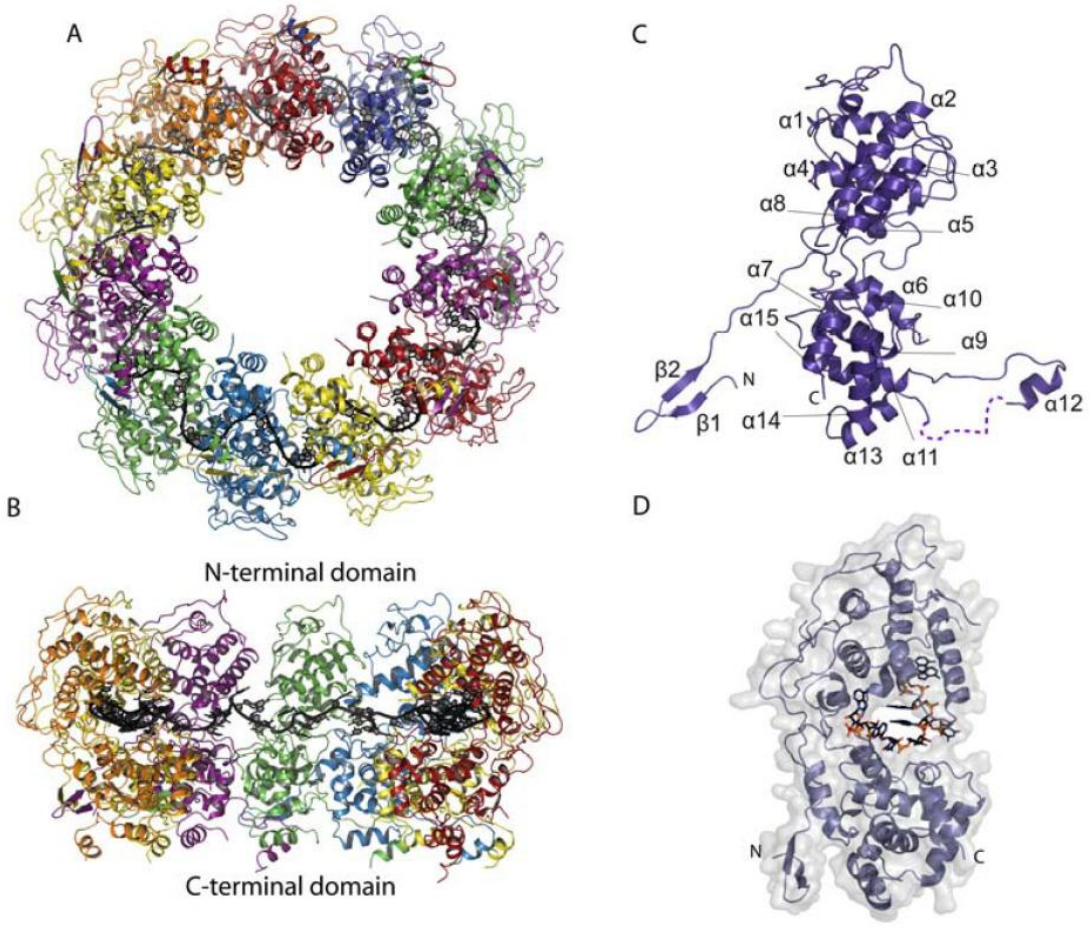


Figure 18 : Structure cristallographique des anneaux N_{11}-ARN du virus de la rage (Albertini, Wernimont et al. 2006)**.** (A) Vue de dessus de l'anneau N_{11}: chaque nucléoprotéine est colorée différemment et l'ARN est schématisé en noir. (B) Vue de coté de l'anneau montrant l'intérieur de six protomères de nucléoprotéines. L'ARN est lié dans la cavité centrale située entre les deux domaines de la nucléoprotéine. (C) Protomère de nucléoprotéine en vue de face indiquant la localisation des différents éléments de structure secondaire. (D) Protomère de nucléoprotéine représenté en cartoons et en surface, permettant d'observer la fixation des 9 bases d'ARN.

Les interactions entre la nucléoprotéine et l'ARN sont principalement d'ordre électrostatique et font intervenir les nombreuses lysines et arginines de la cavité de N et les phosphates de l'ARN (Albertini, Wernimont et al. 2006; Green, Zhang et al. 2006; Luo, Green et al. 2007; Albertini, Schoehn et al. 2008), en accord avec l'hypothèse formulée par Iséni (Iseni, Baudin et al. 2000). Chaque nucléotide est engagé dans des contacts polaires *via* ses groupements phosphates (7 des 9 nucléotides (nt)) ou son ribose (2 des 9 nt). Les trois

premières bases présentent un empilement des orbitales π (« π-stacking ») avec les deux dernières bases liées au protomère de nucléoprotéine précédent. Les nucléotides 4, 6 et 7 sont également empilés, et leurs bases pointent vers l'intérieur de la cavité. La base du nucléotide 5 pointe quant à elle vers le solvant, tout comme les bases 1, 2, 3, 8 et 9 (Figure 19).

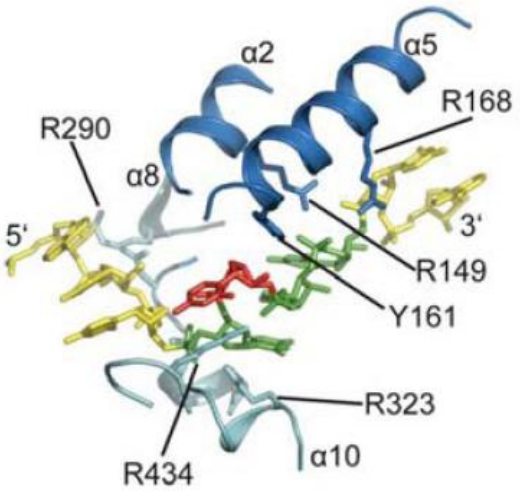


Figure 19: Interaction entre la nucléoprotéine et l'ARN (Albertini, Wernimont et al. 2006). L'ARN est schématisé en bâtonnets. La région NTD est colorée en bleu foncé et le CTD en bleu clair. Les nucléotides dont les bases sont accessibles au solvant sont indiqués en jaune et les bases qui font face à la protéine en vert. Le nucléotide central à la pointe de l'hélice gauche formé par le segment d'ARN est en rouge. Les résidus de N qui contactent les phosphates et les riboses du segment d'ARN sont schématisés en bâtonnets.

Les contacts entre protomères sont médiés par une large surface d'interaction entre les CTDs (~ 2700 $Å^2$), et par deux sous-domaines qui émergent du protomère en N et C-terminal et forment des contacts spécifiques entre les protomères M_{-1} et M_{+1} (Figure 20), assurant ainsi la stabilité de la polymérisation. Les structures de type anneau sont extrêmement rigide, cependant la présence de ces domaines d'échange entre protomères permet de comprendre l'origine moléculaire de la flexibilité observée au niveau des paramètres hélicoïdaux pour les nucléocapsides. En effet, il est probable que la présence de ces contacts spécifiques permette des variations relativement importantes d'angle ente les sous-unités, tout en maintenant les contacts entre les sous-domaines et donc la stabilité de la structure.

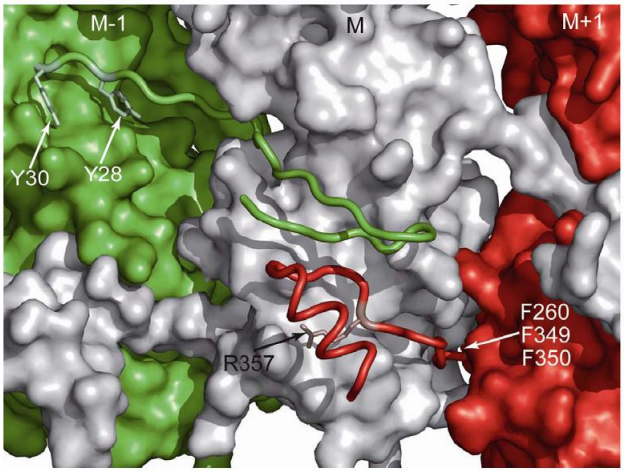


Figure 20: Régions charnières responsables de la polymérisation de la nucléoprotéine (Albertini, Wernimont et al. 2006)**.** Modèle en représentation de surface de l'anneau N_{11}, vue de coté. Chaque monomère est coloré différemment. Le sous-domaine N-terminal du protomère M atteint le protomère M_{+1}, et le sous-domaine C-terminal du protomère M contacte le protomère M_{-1}. Cet arrangement aboutit à l'interaction de la région charnière du NTD M_{-1} et du CTD M_{+1} à la surface du protomère M.

Les nucléoprotéines du virus de la rage et du VSV possèdent des structures remarquablement similaires (Figure 21), permettant un alignement sur base des éléments de structure secondaire (Luo, Green et al. 2007; Albertini, Schoehn et al. 2008), et ce en dépit d'une homologie faible au niveau des séquences (15% d'identité). Les résidus basiques responsables de l'interaction avec l'ARN sont très conservés entre les nucléoprotéines du virus de la rage et de la stomatite vésiculaire, de même que l'agencement tertiaire des hélices α composant les domaines N- et C-terminaux (Figure 21). La différence principale entre ces deux structures réside dans la longueur des boucles C-terminales désordonnées, impliquées dans la fixation de la phosphoprotéine (Schoehn, Iseni et al. 2001), qui contiennent par ailleurs le site de phosphorylation Ser389 chez le virus de la rage (Wu, Lei et al. 2003), absent chez le VSV.

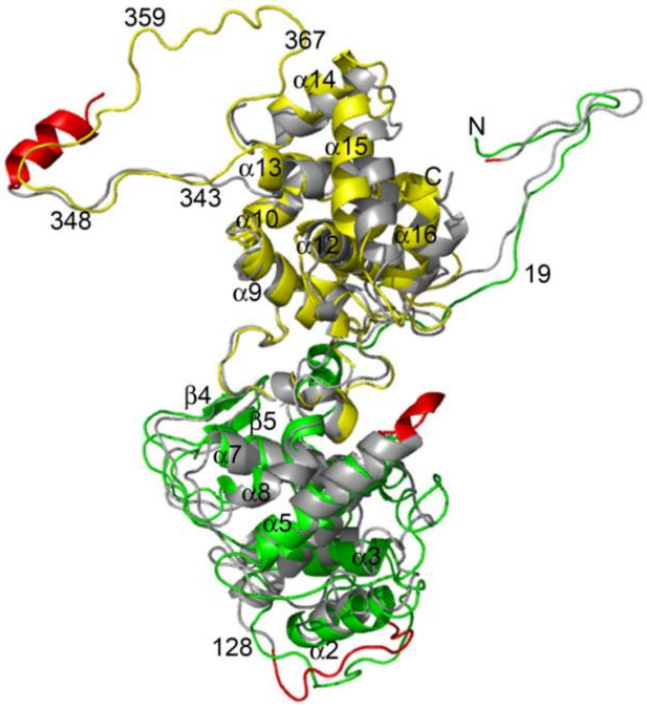


Figure 21: Comparaison des structures des nucléoprotéines du virus de la rage et du VSV (Luo, Green et al. 2007)**.** Les structures des nucléoprotéines du virus de la rage et du VSV sont superposées (*rmsd* 2,18 Å) et représentées en cartoons. Le domaine N-terminal de VSV N est coloré en vert et le domaine C-terminal est en jaune. La nucléoprotéine du virus de la rage est représentée en gris, et les régions présentant des différences structurales importantes sont en rouge. Les résidus N- et C-terminaux, ainsi que les éléments de structure secondaire sont indiqués (d'après la numérotation de la protéine N du VSV).

A l'heure actuelle, six structures de nucléoprotéine sont disponibles: deux nucléoprotéines de *Rhabdoviridae* (Figure 22 A), dans lesquelles chaque protomère est en interaction avec 9 nucléotides d'ARN simple brin (Albertini, Wernimont et al. 2006; Green, Zhang et al. 2006), la nucléoprotéine du virus respiratoire syncytial (Figure 22 B) (*Paramyxoviridae*) (Tawar, Duquerroy et al. 2009) qui révèle un mode d'interaction avec l'ARN radicalement différent de ce qui est observé chez les *Rhabdoviridae*, la nucléoprotéine du virus de la maladie de Borna (Figure 22 C) (*Bornaviridae*), cristallisée en l'absence d'ARN sous la forme d'homo-tétramères (Rudolph, Kraus et al. 2003), la nucléoprotéine du virus de la grippe (Figure 22 D) (*Orthomyxoviridae*), également cristallisée sans ARN, sous

la forme d'homo-trimères (Ye, Krug et al. 2006) et la nucléoprotéine du virus de la fièvre de la vallée du Rift (*Bunyaviridae*) (Raymond, Piper et al. 2010). Cependant, le repliement protéique de cette dernière structure pourrait avoir été affecté par le processus de renaturation employé (Ferron F., communication personnelle).

Même s'il est difficile de superposer les structures de nucléoprotéines des *Rhabdoviridae* avec celles des autres virus à ARN négatif, on observe la présence de deux domaines (NTD et CTD), principalement composés d'hélices α, et responsables de l'interaction avec l'ARN (Figure 22 A, B & C). Cependant, on ne retrouve pas cette séparation en deux domaines pour le virus de la grippe, dont le génome est segmenté (Figure 22 D). La nucléoprotéine du virus de la grippe fixe 24 nucléotides, et l'ARN est facilement éliminé par des traitements à haute molarité en sel. Aussi, chez ce virus, les bases sont plus exposées au solvant (Baudin, Bach et al. 1994) que chez les *Rhabdoviridae* (Iseni, Baudin et al. 2000)et les *Paramyxoviridae* (Iseni, Baudin et al. 2002).

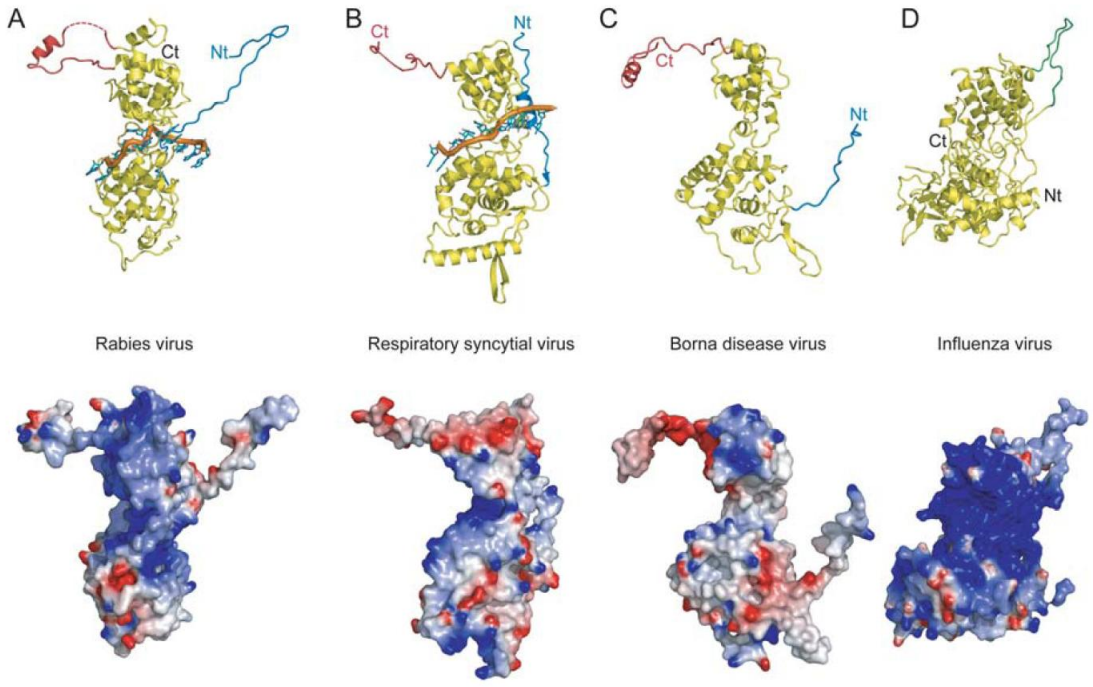


Figure 22: Comparaison des structures de nucléoprotéines de différents virus à ARN négatif (Ruigrok and Crépin 2010)**.** (A) Nucléoprotéine du virus de la rage liée à l'ARN (code PDB : 2GTT). Par souci de clarté, 11 nucléotides sont représentés, bien que chaque protomère n'en fixe que 9. (B) Nucléoprotéine du virus respiratoire syncytial (code PDB: 2WJ8). Ici encore, 9 nucléotides sont représentés, bien que chaque protomère n'en fixe que 7. (C) Nucléoprotéine du virus de la maladie de Borna (code PDB : 1N93). (D) Nucléoprotéine du virus de la grippe (code PDB : 2QO6). Cette protéine ne possède pas de domaines N- ou C-terminaux. L'épingle β impliquée dans l'interaction entre monomères est indiquée en vert. En haut, les structures sont représentées en cartoons jaunes, les sous-domaines N- et C-terminaux impliqués dans la polymérisation sont colorés en bleu et en rouge, respectivement. Les extrémités N- et C-terminales sont également indiquées. En bas, les protéines sont représentées en surfaces électrostatiques colorées du bleu (+5kT/e) au rouge (-5kT/e). L'ARN présent dans la structures du complexe N-ARN du virus de la rage, ainsi que dans celle du virus respiratoire syncytial ont été enlevés, de manière à visualiser le potentiel électrostatique fortement positif de la cavité de fixation à l'ARN.

- *La phosphoprotéine (P)*

Tous les *Mononegavirales* possèdent une phosphoprotéine (P). Les séquences, tailles et degrés d'oligomérisation des phosphoprotéines de ces différents virus sont très variables (568 résidus pour le virus Sendai, 297 pour le virus de la rage et 265 résidus pour le virus de la stomatite vésiculaire).

Cependant, les phosphoprotéines ont des caractéristiques communes :

- elles **possèdent des sites de phosphorylation** permettant de moduler leur activité (Das, Gupta et al. 1995; Das, Schuster et al. 1995; Byrappa, Pan et al. 1996; Spadafora, Canter et al. 1996; Chen, Das et al. 1997; Huntley, De et al. 1997; Mathur, Das et al. 1997; Pattnaik, Hwang et al. 1997; Schwemmle, De et al. 1997; Hwang, Englund et al. 1999; Raha, Chattopadhyay et al. 1999; Gupta, Blondel et al. 2000; Raha, Samal et al. 2000; Das and Pattnaik 2004; Toriumi, Eriguchi et al. 2004; Moseley, Filmer et al. 2007).

- elles jouent un rôle de **chaperonne pour la nucléoprotéine** empêchant ainsi sa liaison aux ARN cellulaires (Masters and Banerjee 1988; Takacs, Perrine et al. 1991; Takacs, Das et al. 1993; Curran, Marq et al. 1995; Takacs and Banerjee 1995; Gupta and Banerjee 1997; Mavrakis, Iseni et al. 2003; Mavrakis, Mehouas et al. 2006; Chen, Ogino et al. 2007).

- elles se fixent sur le **complexe nucléoprotéine-ARN** (Paul, Chattopadhyay et al. 1988; Takacs, Das et al. 1993; Takacs and Banerjee 1995; Schoehn, Iseni et al. 2001; Ribeiro, Favier et al. 2008)**.**

- elles constituent le **cofacteur de l'ARN polymérase virale ARN-dépendante** (Emerson and Yu 1975; Mellon and Emerson 1978; De and Banerjee 1984; Thornton, De et al. 1984; De and Banerjee 1985; Gill, Chattopadhyay et al. 1986; Emerson and Schubert 1987; Howard and Wertz 1989; Green, Macpherson et al. 2000)**.**

- Elles forment des **homo-oligomères** (Tarbouriech, Curran et al. 2000; Tarbouriech, Curran et al. 2000; Choudhary, Malur et al. 2002; Ding, Green et al. 2006; Llorente, Garcia-Barreno et al. 2006; Cevik, Smallwood et al. 2007; Gérard, Ribeiro et al. 2007; Ivanov, Crepin et al. 2010).

- Elles **interfèrent avec les mécanismes d'immunité innée** de la cellule hôte et jouent un rôle dans l'infectivité du virus (Blondel, Regad et al. 2002; Brzozka, Finke et al. 2005; Vidy, Chelbi-Alix et al. 2005; Brzozka, Finke et al. 2006; Chelbi-Alix, Vidy et al. 2006; Vidy, Bougrini et al. 2007; Ito, Moseley et al. 2010).

- Elles **interagissent avec le cytosquelette** (Fujinami, Oldstone et al. 1983; Zafrullah, Ozdener et al. 1997; Jacob, Badrane et al. 2000; Raux, Flamand et al. 2000; Poisson, Real et al. 2001; Moseley, Lahaye et al. 2009).

REVIEW 1: Structure and Functions of Rabies Virus Phosphoprotein

Auteurs: Cédric Leyrat, Euripedes de Almeida Ribeiro Jr, Francine Gérard, Ivan Ivanov, Rob Ruigrok et Marc Jamin.

Soumis à *Future Virology*

Structure and functions of rabies virus phosphoprotein

Cédric Leyrat, Euripedes A. Ribeiro Jr., Francine C.A. Gérard, Ivan Ivanov,

Rob W. H. Ruigrok and Marc Jamin

UMI 3265 UJF-EMBL-CNRS, Unit of Virus Host Cell Interactions
6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

Summary

Rabies is an incurable albeit preventable disease that remains an important human health issue, still killing more than 50,000 people each year. Its causative agent, the rabies virus, is a negative-sense RNA virus, whose genome encodes five proteins. Three of these proteins, the nucleoprotein (N), the phosphoprotein (P) and the large protein (L), are required to synthesize viral RNA in an efficient and regulated manner. P plays multiple roles during the transcription and replication of the RNA genome. It acts as a non-catalytic cofactor of the L polymerase and it chaperones N. Recent structural characterizations of RV P revealed that P forms elongated and flexible dimers, and uncovered the structural basis of its modular organization, revealing the existence of two independent structured domains and two long intrinsically disordered regions (IDRs). In addition, recent studies also revealed that P interacts with nucleo-cytoplasmic trafficking carriers and with the host cell cytoskeleton, probably allowing viral components to be transported within the host cell, and it blocks the innate immune response by inhibiting different steps of the interferon pathway. With multiple binding sites for different viral and cellular partners located in either its structured or disordered regions, P appears as a flexible "hub" protein, that connects viral or cellular proteins and allows their assembly into multi-molecular complexes. These new findings shed light on the mechanism of replication of the virus and on the intimate interactions between the virus and its host cell and will also help to identify new targets for the development of antiviral treatments.

Keywords: Virus, rabies, replication, phosphoprotein, intrinsically disordered region

Rabies and the family *Rhabdoviridae*

Rabies.

Rabies is a multihost zoonotic disease, whose transmission by dog bite and fatal issue have been known since the 23rd century BC (Théodoridès 1986). The main causative agent of rabies is the rabies virus (RV, genotype 1), the prototypic member of the genus *Lyssavirus*, although members of several other genotypes can also elicit the disease. RV is a neurotropic virus that causes an encephalitis with a complex pathogenesis leading to different forms of the disease ("furious" or encephalitic, "dumb" or paralytic, non-classic rabies) (Hemachudha, Laothamatas et al. 2002; Dietzschold, Li et al. 2008). The encephalitis is incurable and fatal in 100 % of the cases but the infection can be prevented. Since the first successful use of rabies vaccine by Louis Pasteur in 1885, the disease is under control in Western Europe and North America by the application of pre-exposure and post-exposure treatments (Rupprecht, Willoughby et al. 2006). However, rabies remains an important, albeit neglected, human health problem, killing more than 50,000 people each year worldwide (one person every 10 minutes) mainly in Asia and Africa (see WHO Fact Sheet N° 99, 2008) and imposing important economic burden in affected populations (Knobel, Cleaveland et al. 2005). Moreover, recent reports

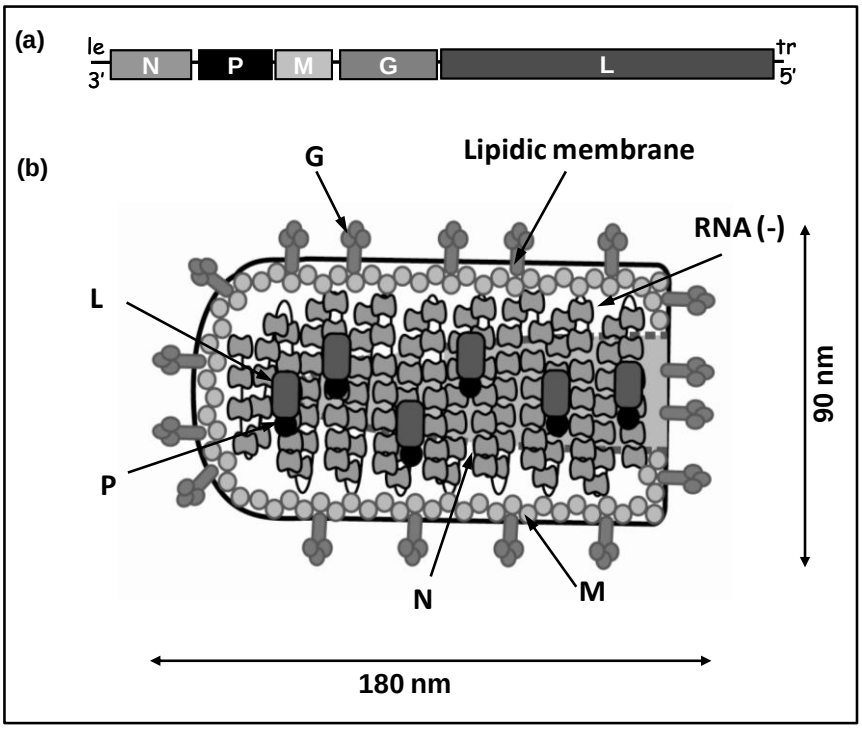


Figure 1. The rabies virus. (a) RV genome. The genome encodes five proteins and contains untranslated 3' leader (le) and 5' trailer (tr) sequences. **(b) Structure of the virion.** Schematic representation of the bullet-shaped RV showing the dimensions of the particle and the relative organization of the glycoprotein (G), matrix protein (M), nucleoprotein-RNA complex (N-RNA), phosphoprotein (P) and large subunit of the RNA-dependent RNA polymerase (L). The L and P proteins are localized inside the N-RNA complex but are shown here outside for presentation purpose.

suggest that the number of cases is substantially underestimated in poor rural area (Cleaveland, Fevre et al. 2002). The main natural reservoirs of rabies virus are in the orders *Carnivora* (dog, fox, raccoon, skunk) and *Chiroptera* (bats). The mechanism of cross-species transmission remains poorly understood (Streicker, Turmelle et al. 2010), but new strains, potentially unsusceptible to vaccination against classical rabies (Hanlon, Kuzmin et al. 2005; Weyer, Kuzmin et al. 2008) could emerge and spread again in human population (Fooks 2005; Leslie, Messenger et al. 2006).

The Rhabdoviridae.

Rabies virus (RV) is a negative RNA virus that belongs to the family *Rhabdoviridae*. This large family of RNA viruses includes numerous pathogens of plants and animals, which vary widely in terms of cellular interactions and replicate either in the cytoplasm or in the nucleus of the host cell. The vesicular stomatitis virus (VSV) in the genus *Vesiculovirus* is the other prominent member of the *Rhabdoviridae*. It is an arthropod-borne virus that causes a mild disease in horses, cattle and pigs and rarely transmits to human. However, it has an important economical impact in the farm industry in the Americas where the virus is endemic. For many years, VSV served as a model system for studying the replication cycle of the Rhadoviruses and, more generally, the interactions between viruses and their host cells. More recently, VSV has been used as a vaccine vector and as an oncolytic virus (Lichty, Power et al. 2004; Rose, Publicover et al. 2008). Other *Vesiculoviruses* such as Chandipura virus or Piry virus infect humans. In recent outbreaks in India, Chandipura virus produced acute encephalitis with high fatality ratio. Finally, numerous Rhabdoviruses infect fishes or plants and cause economic damage in industrial fisheries and agricultural farms. The family *Rhabdoviridae* is grouped in the order *Mononegavirales* (*MNV*) with the *Filoviridae* (Ebola and Marburg viruses), the *Paramyxoviridae* (measles, mumps, respiratory syncytial viruses) and the *Bornaviridae* (Borna disease virus) because all these viruses share similar genome and structural organizations as well as similar strategies of RNA replication and transcription (Pringle 1997).

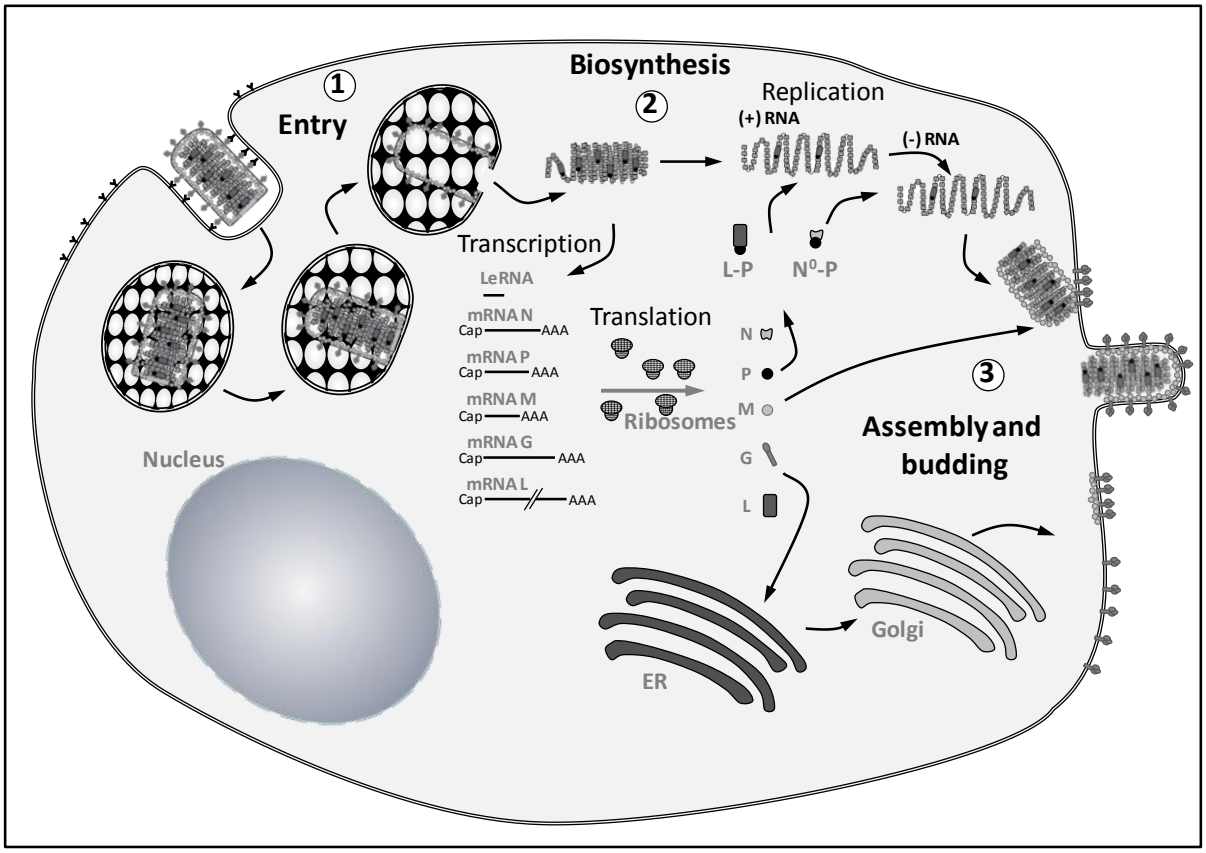


Figure 2. Rabies Virus replication cycle. The replication of rabies virus occurs in three main phases. The first phase (1) includes the attachment of the viral particle onto the surface of the host cell, followed by its internalization into an endosome and its intra-cellular axonal transport towards the cell body. These steps depend on the G protein. Once the viral particle has reached the neuronal cell body, the NC is released in the cytoplasm through fusion of the viral membrane with the endosomal membrane. The fusion is induced by a pH dependent conformational change of the G protein (Gaudin, Tuffereau et al. 1999). The recent crystal structures of the G protein of VSV in pre- and post-fusion conformations have provided a glimpse at the conformational changes of G that underlie the fusion process (Roche, Bressanelli et al. 2006; Roche, Rey et al. 2007). In the second phase (2), the viral genome is transcribed and replicated. Both processes occur entirely in the cytoplasm, in inclusion bodies named Negri bodies, that are typical of rabies virus-infected neurons (Lahaye, Vidy et al. 2009). Only three of the viral proteins, N, P and L, are required to synthesize viral RNA in an efficient and regulated manner (Barr, Whelan et al. 2002). In the early times, transcription of the incoming viral genome takes place in a sequential process, producing a positive-stranded leRNA and five capped and poly-adenylated mRNAs. The abundance of the mRNAs decreases with distance from the 3' end of the genome because the polymerase initiates transcription at a single entry site and transcribes the genes in order from the 3' to the 5' ends. The viral proteins are then produced by the host cell translation machineries. In contrast with VSV infection, in which the expression of host cell genes is completely shut-off, allowing a rapid and efficient production of viral proteins, RV infection maintains the expression of host cell genes and preserves cellular functions required for the correct spread of the virus. In rabies virus infection, the M protein binds to eIF3h and down-regulates the canonical cap-dependent translation initiation process (Komarova, Real et al. 2007). Because the cap structure of *Rhabdoviridae* is different from the cellular cap structure (Ogino and Banerjee 2007), it is probable that expression of RV proteins is less affected. Later, the viral polymerase switches from transcription to replication of the RNA genome and starts producing a full-length positive-sense anti-genome that subsequently serves as a template for the synthesis of new viral RNA genomes. In VSV, two multi-molecular complexes that contain the viral L, P and N proteins, but differ in their content in host-cell proteins, have been postulated to separately catalyze transcription and replication (Qanungo, Shaji et al. 2004). The mRNAs are not encapsidated by N (Patton, Davis et al. 1984), but both antigenomes and newly synthesized genomes are encapsidated by N, and therefore replication requires the continuous production of soluble N. The switch from transcription to replication could be regulated by the accumulation of soluble N proteins, but it also depends on the M protein (Finke and Conzelmann 2003), and the precise mechanism of this regulation remains unknown. In the third phase (3), the viral components migrate to the host cell plasma membrane where they assemble. New viral particles bud out, taking away a piece of the host cell membrane containing the G protein. The M protein interacts with both the N protein and host membranes (Dancho, McKenzie et al. 2009) and can therefore serves as scaffold for assembling the viral components at the inner face of the cellular membrane. The membrane must then deform and pinch off in a membrane fission step. The rabies virus M protein contains late domains involved in recruiting cellular machineries involved in the formation of vesicles that bud away from the cytoplasm, such as the proteins involved in the vacuolar protein sorting (VPS) process. In particular, binding of M to Nedd4, an E3 ubiquitin ligase that recruits proteins in the VPS pathway, is required for efficient budding of rabies virus (Harty, Paragas et al. 1999; Wirblich, Tan et al. 2008).

The rabies virus and its life cycle

The RNA genome.

The viral genome of RV is made of a single molecule of negative sense RNA of about 12 kb, flanked on both sides by untranslated leader (le) and trailer (tr) RNA regions (Figure 1a). It comprises only five genes that are common to all members of the *MNV* and that encode successively from the 3' terminus, the nucleoprotein (N), the phosphoprotein (P), the matrix protein (M), the glycoprotein (G) and the large subunit of the RNA-dependent RNA polymerase (L) (Figure 1a).

Structure of the viral particle.

RV is a bullet-shaped particle of 180 nm in length and 90 nm in diameter (Schoehn, Iseni et al. 2001) (Figure 1b). The cryo-electron microscopy (EM) reconstruction of the VSV particle, revealed the molecular organization of the viral proteins in three concentric layers (Ge, Tsao et al. 2010). The outer shell is a host cell-derived lipid membrane decorated with trimeric G spikes. The middle layer is formed by M proteins which are organized in a regular helix onto which the trimeric G spikes attach. The inner layer is the nucleocapsid (NC), the actual infectious part of the virus, composed of the genomic RNA molecule associated with the N, P and L proteins. N has two domains forming a groove into which the RNA molecule binds through interactions with the sugar-phosphate backbone (Iseni, Baudin et al. 2000; Schoehn, Iseni et al. 2001; Albertini, Wernimont et al. 2006). N proteins bind every 9 nucleotides and completely enwrap the genomic RNA, forming a long N-RNA complex organized in a regular helix (Schoehn, Iseni et al. 2001). P and L proteins attach to this N-RNA complex but cannot be visualized in the cryo-EM reconstruction because they are not organized in a highly symmetrical fashion.

The viral life cycle.

All intricacies of the rabies life cycle have not yet been unravelled, and recent findings have been reviewed (Schnell, McGettigan et al. 2010). Figure 2 summarizes the main events of the viral life cycle, which is divided in three main phases. Numerous features of the cellular replication process are common to RV and VSV, although RV replicates more slowly and is less cytopathic than VSV, indicating different regulatory mechanisms and different modes of interference with the host cell.

Structure and functions of the phosphoprotein

As its name indicates, RV P is a phosphorylated protein. It is a multifunctional protein that plays multiple roles in the viral transcription and replication processes. Firstly, P plays an essential role in the encapsidation of both genomic and antigenomic RNA molecules. It chaperones nascent N molecules by forming a complex named N^0-P (the subscript ^{0} indicates the absence of RNA). Secondly, P associates with L to form an active RNA-dependent RNA polymerase (L-P complex), which uses the N-RNA complex rather than the naked RNA as a functional template.. Host cell co-factors are also found in association with the polymerase complex, but their roles and mechanism of action have not yet been determined (Qanungo, Shaji et al. 2004). In addition, P meddles with the host cell intracellular trafficking systems and acts at several stages of the interferon pathway for inhibiting the cellular innate immune response. The recent structural characterizations of both RV and VSV P, which are discussed in this review, bring new lights on the structural organization of these proteins and on the way they achieve their multiple functions. Data on VSV P are included since this protein shares many functional and structural features with RV P.

P is a heterogeneous protein

Truncated forms of P.

In infected cells, RV P is present in several truncated forms and in different phosphorylation states (Figure 3). Apart from the full-length P, four N-terminally truncated forms of RV P (P_2: aa 20-297, P_3: aa 53-297, P_4: aa 69-297 and P_5: aa 83-297) are produced from internal in-frame AUG initiation codons by a ribosomal leaky scanning mechanism (Figure 3) (Chenik, Chebli et al. 1995). These forms of P have different subcellular localizations resulting from the interplay between nuclear import and export processes mediated by a C-terminal nuclear localization signal (NLS) and nuclear export signals (NES) in the N- and C-terminal part of P

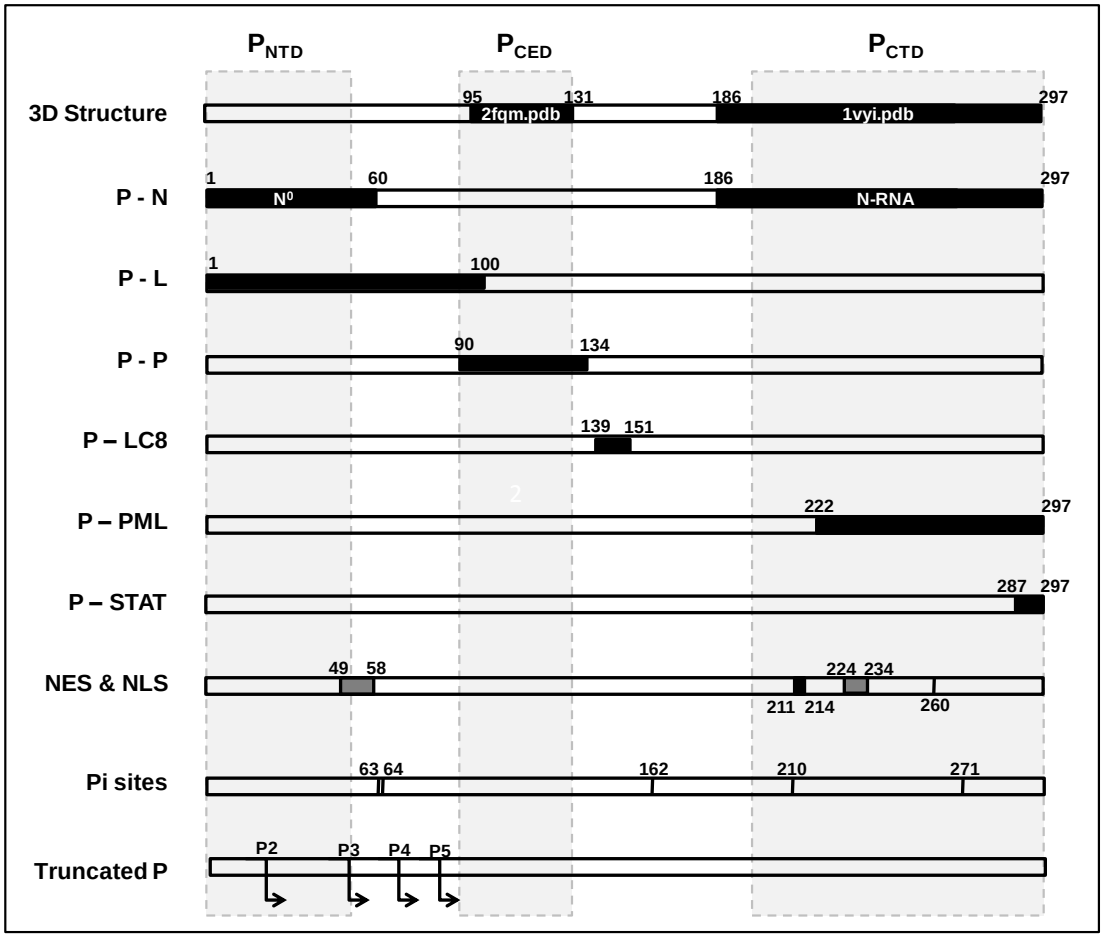


Figure 3. Functional and structural domains of RV P. The black and gray boxes represent the localization along the amino acid sequence of the known 3D structures and of the different functional regions. Numbers indicates the defined boundaries. The three shaded area correspond to the localization of structured domains predicted from the amino acid sequence by a recent meta-prediction (Gérard, Ribeiro et al. 2009).

(Pasdeloup, Poisson et al. 2005; Moseley, Filmer et al. 2007; Vidy, El Bougrini et al. 2007) (see below). Full-length P and P_2 are found in the cytoplasm, whereas P_3, P_4 and P_5 are found in the nucleus. The exact biological role of the different forms of RV P has yet to be determined. In the cytoplasm, full-length P participates in the transcription and replication processes, whereas in the nucleus P3 interferes with the interferon response pathway (Vidy, Chelbi-Alix et al. 2005; Vidy, El Bougrini et al. 2007; Ito, Moseley et al. 2010), but the functions of P2, P4 and P5 remain to be characterized. In VSV, as in other members of the *MNV*, truncated forms of P are also produced from the same mRNA, some of which blocks the innate immune response (Childs, Andrejeva et al. 2009).

Phosphorylation of P.

RV P and VSV P are phosphorylated in the virion and in infected cells, and numerous studies, in particular for VSV, have been devoted to deciphering the role of these phosphorylations. In both RV and VSV, phosphorylation occurs in two regions of P, in the N-terminal acidic region and in the C-terminal domain (Figure 3). In both viruses, phosphorylation is mediated by two different kinases and occurs sequentially, resulting in the existence of two main forms of the protein (Barik and Banerjee 1992; Gupta, Blondel et al. 2000). Phosphorylation of VSV P regulates transcription and replication, although the precise mechanism of these regulations is unknown. Phosphorylation in the N-terminal acidic region by casein kinase II is required for transcription, while the subsequent phosphorylation in the C-terminal domain by a kinase associated with the L protein regulates the binding of the L/P polymerase complex to its N-RNA template is thus required for RNA synthesis (Chattopadhyay and Banerjee 1987; Hwang, Englund et al. 1999). It has been proposed that

phosphorylation in the N-terminal region induces the self-assembly of P and thereby regulates the association with L and the activation of transcription (Das, Gupta et al. 1995; Gao and Lenard 1995; Gao and Lenard 1995; Pattnaik, Hwang et al. 1997). However, a double mutation (S60D/T62D) that mimicks phosphorylation and renders P constitutively active in transcription (Gao and Lenard 1995), has no effect on the oligomerization of P (Gérard, Ribeiro et al. 2007). In RV, protein kinase C (PKC) phosphorylates residues in the C-terminal part of P (S162, S210 and S271), whereas a rabies virus-specific kinase phosphorylates S63 (Figure 3) (Gupta, Blondel et al. 2000). Phosphorylation in the N-terminal region of RV P has no effect on the oligomerization of the protein and no known function (Gigant, Iseni et al. 2000), whereas phosphorylation in P_{CTD}, in particular of S210, appears to modulate the nucleocytoplasmic distribution of P and plays a role in blocking the innate immune response (Moseley, Filmer et al. 2007) (see below).

RV P has a modular organization and contains long intrinsically disordered regions

Functional modules.

The existence of independent functional modules in P was first revealed by deletions and mutations in VSV P. Early studies identified three distinct functional regions directly involved in transcription and showing a high conservation rate among VSV strains (Gill, Chattopadhyay et al. 1986; Emerson and Schubert 1987; Das and Pattnaik 2005). A first N-terminal region (aa 1 – 137) is highly acidic and contains binding sites for L (Emerson and Schubert 1987) and N^0 (Takacs, Das et al. 1993). It also contains three phosphorylation sites involved in transcriptional regulation (Barik and Banerjee 1992; Barik and Banerjee 1992). The second and third functional regions form the C-terminal part of P (aa 211 – 240 and aa 245-265). The second region binds to L and contains phosphorylation sites that control replication (Hwang, Englund et al. 1999), while the third region (aa 245 – 265) is basic and contains the binding site for the N-RNA template (Paul, Chattopadhyay et al. 1988; Das, Pattnaik et al. 1997). The remaining region (aa 138 – 210), named the "hinge" region, is hypervariable and its distal part (aa 201-220) plays a role in the assembly of infectious particles (Das and Pattnaik 2005). A central region of VSV P, at the junction of the acidic N-terminal region and of the "hinge" region contains a self-association domain (Ding, Green et al. 2006). In RV P, similar functions were mapped to corresponding N-terminal and C-terminal regions (Gigant, Iseni et al. 2000; Schoehn, Iseni et al. 2001; Mavrakis, McCarthy et al. 2004; Mavrakis, Mehouas et al. 2006; Ribeiro, Favier et al. 2008; Gérard, Ribeiro et al. 2009), and a self-association domain was mapped in the central part of the molecule (Jacob, Real et al. 2001; Gérard, Ribeiro et al. 2009; Ivanov, Crepin et al. 2010) (Figure 3), suggesting a common modular organization. Comparison with the P protein of Sendai virus and measles virus of the *Paramyxoviridae* and amino acid sequence analysis suggested that RV P contains structured domains separated by intrinsically disordered regions (IDRs) (Karlin, Ferron et al. 2003; Gérard, Ribeiro et al. 2009).

Intrinsically disordered regions (IDRs).

Since the early days of protein folding studies, biological activities have been associated with the adoption by the polypeptide chain of a well-defined three-dimensional structure. However, this structure-function paradigm was recently challenged with the discovery that many proteins are intrinsically disordered or contain intrinsically disordered regions (IDRs) under physiological conditions (Wright and Dyson 1999). Various algorithms were developed for identifying and locating IDRs in proteins (He, Wang et al. 2009), and complete genome surveys revealed that as much as one third of the proteins from eukaryotes contain long IDRs (Tokuriki, Oldfield et al. 2009). Even more surprisingly, disorder appears to be required for the biological function of these proteins (Tompa 2002). IDRs are found in numerous proteins involved in molecular recognition, cell signalling and regulation control (Uversky, Oldfield et al. 2005) and many of them serve as flexible scaffold for the assembly of multi-molecular complexes (Spolar and Record 1994; Wright and Dyson 2009). IDRs exhibit different modes of binding to their partners, some fold upon binding whereas others remain partially disordered, and thereby can have a wide range of effects on the energetics of the interactions.

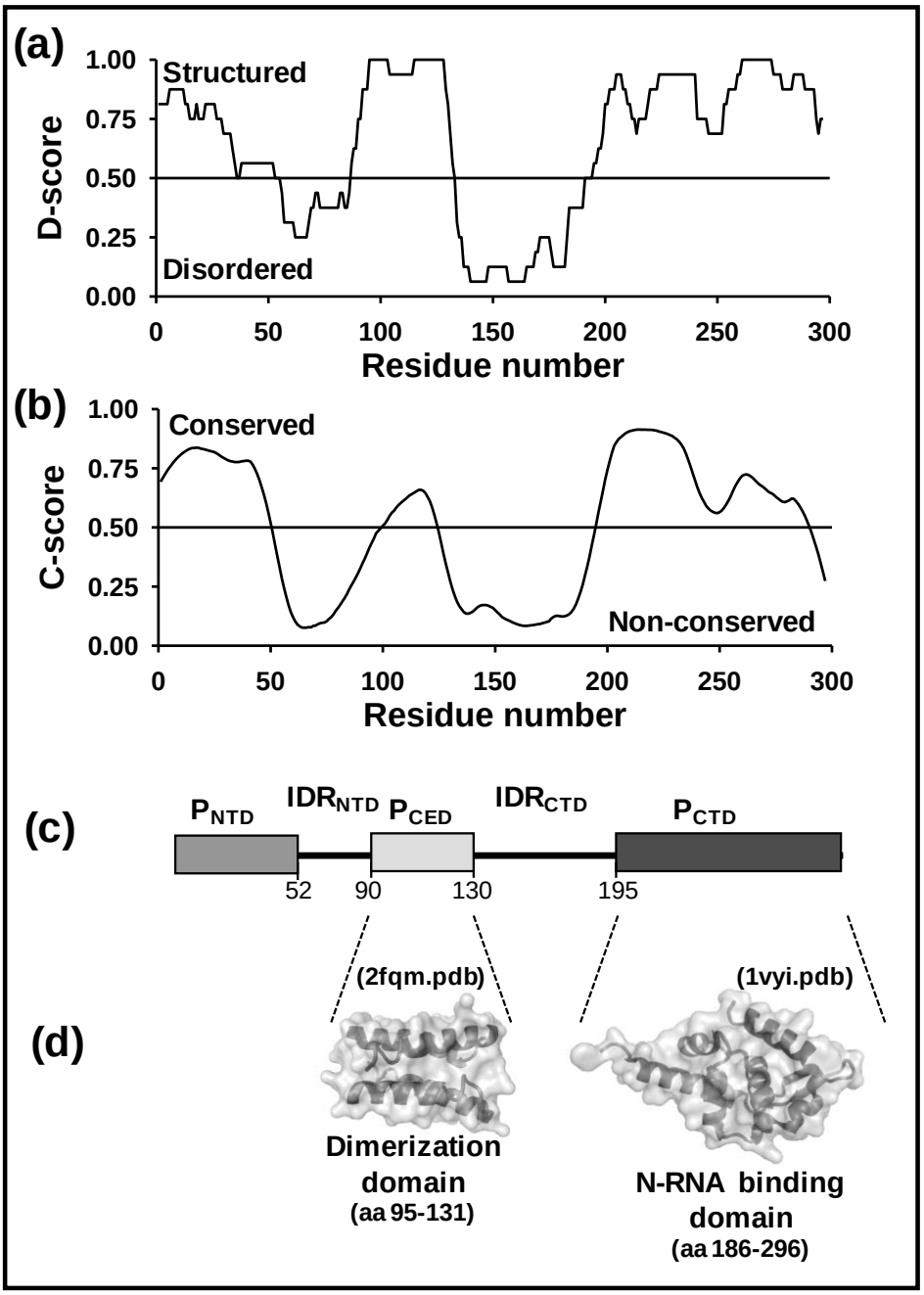


Figure 4. Modular organization of RV P. (a) Metaprediction of IDRs. A consensus prediction for the localization of IDRs was obtained by combining results from 16 independent predictors available through WEB servers (Gérard, Ribeiro et al. 2009). For each disorder prediction, residues were simply ranked in a binary manner as ordered or disordered using the default threshold set for each algorithm. For each predicting algorithm, a score of 0 was attributed to a residue when it was predicted to be disordered, whereas a score of 1 was attributed when it was predicted to be in a structured region. The D-score was calculated by adding the values for each residue and dividing by the number of used algorithms. We arbitrarily defined a threshold at 0.50, above which residues are assigned as structured, and which residues are assigned as disordered. **(b) Sequence conservation.** The amino acid sequences of the P protein from 13 *Lyssavirus* were aligned and the conservation rate for each amino acid was scored by using the AL2CO software (Pei and Grishin 2001). A C-score above 0.5 indicate a conserved position. This analysis revealed the presence in the sequence of three conserved regions (aa 1-50, 101-126 and 194-292) that are located within or correspond to the otherwise predicted structured regions. (**c) Consensus localization of structured and disordered regions.** The consensus disordered regions determined from the D-score suggests the presence of two disordered regions (aa 53-89 and 131-194) and three structured domains (aa: 1- 52, 90-130 and 195-297). (d) Structured domains. The high resolution structures of two domains corresponding closely to the predicted P_{CED} and P_{CTD} have been determined by X-ray crystallography (Mavrakis, McCarthy et al. 2004; Ivanov, Crepin et al. 2010).

Prediction of disordered regions.

A recent meta-prediction of IDRs in RV P hinted at a modular structural organization of RV P with three folded domains, a N-terminal (P_{NTD}), a central (P_{CED}) and a C-terminal domain (P_{CTD}), separated by two IDRs (IDR_{NTD} and IDR_{CTD}) (Figures 4a and 4c) (Gérard, Ribeiro et al. 2009). A multiple alignment of the amino acid sequences of P proteins from 13 members of the genus *Lyssavirus* revealed that the three predicted structured domains were more conserved than the predicted IDRs (Gérard, Ribeiro et al. 2009) (Figure 4b). The precise localization of the boundaries between disordered regions and folded domains was confirmed by dissecting the protein and characterizing its fragments (Gérard, Ribeiro et al. 2009). The predicted P_{CTD} had been previously identified as an autonomous folding domain and its structure solved by X-ray crystallography (Figure 4d) (Mavrakis, McCarthy et al. 2004). The predicted P_{CED} also appeared as an autonomous folding domain and its structure was solved recently (Figure 4d) (Ivanov, Crepin et al. 2010). The N-terminal part of RV P is acidic, as in VSV P, and the predicted P_{NTD} appeared globally disordered (Gérard, Ribeiro et al. 2009). A similar analysis of VSV P also predicted the presence of folded domains in the N and C-terminal parts that closely correspond to some previously identified functional regions. The predicted P_{NTD} corresponds to the N^0 binding site. The identified C-terminal domain (aa 194-265), whose structure could be solved by NMR spectroscopy (Ribeiro, Favier et al. 2008) encompasses the second and third functional regions as well as the distal part of the hinge region required for encapsidation of the RNA. Predictions in the central part of VSV P are not as clear and are discussed here below.

RV P is a dimeric protein

Self-assembly of P.

RV P self-assembles both *in vivo* and *in vitro*, and the oligomerization domain perfectly matches with the predicted P_{CED} (Gigant, Iseni et al. 2000; Jacob, Real et al. 2001; Gérard, Ribeiro et al. 2009) (Figures 3 and 4). Non-phosphorylated RV P as well as isolated P_{CED} produced in bacteria form dimers (Gérard, Ribeiro et al. 2007), whereas the protein deleted from the central region ($P_{\Delta 91\text{-}131}$) forms monomers (Gérard, Ribeiro et al. 2009). In the crystal structure, P_{CED} also forms dimers (Figure 4d) (Ivanov et al., 2010). Similarly, unphosphorylated VSV P also forms dimers (Gérard, Ribeiro et al. 2007), and an autonomous folding domain isolated by limited proteolysis (VSV P_{CED}, aa 107 to 170) is dimeric in the crystal structure (Ding, Green et al. 2004; Ding, Green et al. 2006). However, the meta-prediction of IDRs in VSV P failed to identify the crystallized domain and rather predicted a bipartite folded domain in the centre of P (aa 98-105 and aa 136-186) (Gérard, Ribeiro et al. 2009).

Structure of the central dimerization domain, P_{CED}.

A comparison of the dimeric structures of RV and VSV P_{CED} showed that they are so different from each other that it is impossible to determine whether these domains have evolved from a common ancestor (Ivanov, Crepin et al. 2010). The monomer of VSV P_{CED} is made of a central α-helix flanked by two two-stranded β-sheets. In the dimer, the helices from the two monomers pack against each other in a parallel orientation, and on each side of the helical bundle, the N-terminal β-sheet of one monomer forms a four stranded β-sheet with the C-terminal β-sheet of the other monomer (Ding, Green et al. 2006). The monomer of RV P_{CED} is made of two α-helices forming a hairpin, and the dimer is assembled by packing two monomers in a parallel orientation with an angle of about 50º (Ivanov, Crepin et al. 2010).

Why is P dimeric?

By comparison with the mechanism of transcription/replication of Sendai virus (*Paramyxoviridae)*, it was speculated that oligomerization of P allows the protein to move along the N-RNA template by a cartwheeling mechanism (Kolakofsky, Le Mercier et al. 2004). In Sendai virus, P forms tetramers through a central coiled-coil domain that is required for transcription. In RV, the self-assembly domain of P is dispensable for transcription (Jacob, Real et al. 2001), but a recent study revealed that dimerization is necessary for the association of P3 with the microtubules and its nuclear import (Moseley, Roth et al. 2007; Moseley, Lahaye et al. 2009).

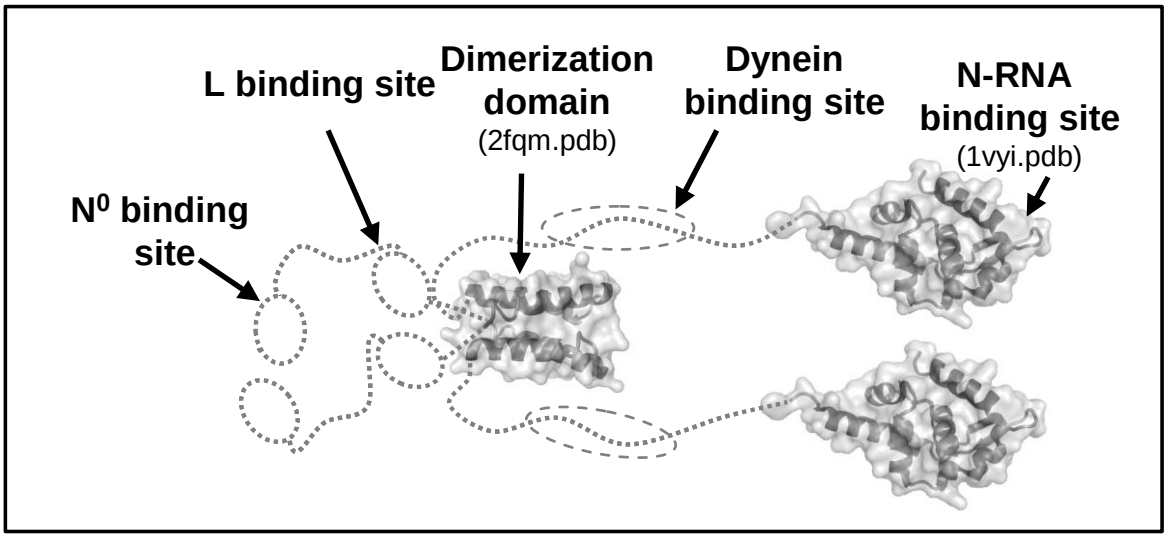


Figure 5. Overall organization of RV P. The known three-dimensional structure of P_{CED} and P_{CTD} are shown as cartoon models and the disordered regions are shown as dotted lines. The dotted circles represent the localization of different binding sites, which could fold upon binding to their partner.

P is a chaperone of nascent nucleoprotein.

The N^0-P complex.

The RNA genome of Rhabdoviruses is protected by the nucleoprotein (N) and, thus, during viral replication, the newly synthesized RNA genomes and antigenomes must be encapsidated by N. N has a strong affinity for RNA and binds non-specifically to cellular RNA. A first role of P is to act as a chaperone of nascent N molecules. P forms with the nucleoprotein (N) a complex named N^0-P complex (the subscript ^{0} denotes the absence of RNA) (Mavrakis, Mehouas et al. 2006; Chen, Ogino et al. 2007), with a stoichiometry of one N plus two P (Mavrakis, Iseni et al. 2003), that maintains nascent N molecules in a soluble and RNA-free form and that is competent for encapsidation of the newly synthesized RNA genomes (Peluso and Moyer 1988; Howard and Wertz 1989).

Binding site for N^0.

Limited proteolysis showed that the N-terminal 40 residues of RV P are necessary and sufficient to maintains the integrity of the soluble N^0-P complex (Mavrakis, Mehouas et al. 2006). Similarly in VSV, the binding site for N^0 is localized in the N-terminal region (aa 11-30) (Chen, Ogino et al. 2007). The binding site for N^0 is thus located within the predicted N-terminal domain, named P_{NTD}. The RNA binding cavity of N has a highly positive electrostatic surface potential (Albertini, Wernimont et al. 2006; Green, Zhang et al. 2006), whereas P_{NTD} is rich in negatively charged residues, suggesting that P_{NTD} binds in the same cavity of N as RNA (Hudson, Condra et al. 1986; Mavrakis, Mehouas et al. 2006).

Although our meta-prediction of IDRs suggested that the first 52 amino acids of RV P form a structured domain (Figure 4a) and that they are highly conserved within the genus *Lyssavirus*, a peptide corresponding to the first 60 residues is globally disordered (Gérard, Ribeiro et al. 2009). Its Stokes's radius measured by size exclusion chromatography is that of a fully unfolded protein of the same molecular mass, and the narrow width of resonances in the NMR spectrum and the poor chemical shift dispersion imply a globally disordered molecule. Similar results were obtained with VSV P_{NTD}, but circular dichroism spectroscopy indicates the presence of low amounts of secondary structure, and NMR relaxation measurements reveal the existence of two transient α-helical elements (aa 2-8 and aa 25-38) (Leyrat et al., *unpublished data*), *i.e.* α-helices that are not stable enough to be present all the time in the protein, but which fold and unfold continuously. This local order in the N-terminal part of P could indicate that P_{NTD} contains a molecular recognition element or MoRE, *i.e.* a short motif within a disordered protein that promotes association with a

partner by folding upon binding (Mohan, Oldfield et al. 2006; Jensen, Houben et al. 2008).

Importance of disorder for the activity of P.

The N-terminal part of RV P (aa 1-92) not only contains the binding site for N^0, but also contains binding sites for L, host cell kinases and importins. IDRs have the ability to interact with multiple structurally diverse partners, and their binding energetics favors exchange between different partners. During the viral replication cycle, the N^0-P complex is not an end-product. It forms only transiently, as an intermediate during the synthesis of new nucleocapsids. The N^0-bound P must be outcompeted by the newly synthesized genomic RNA, and therefore the binding affinity of P for N^0 cannot be too high. The mechanism of folding upon binding provides a means of specific recognition without the corollary of high affinity (Dyson and Wright 2002; Wright and Dyson 2009). Indeed, the folding or the adoption of a rigid structure by the ligand when it binds to its receptor will generally lead to the formation of multiple specific intermolecular interactions that confer a great specificity, and consequently a high affinity. If the ligand is disordered in its unbound form, this strong binding energy is opposed by the high entropy of the disordered protein. We might thus speculate that by being disordered in its free form, P_{NTD} recognizes N^0 with high specificity but moderate affinity, thus allowing its displacement by the newly synthesized viral RNA, but also by host cell proteins.

P is a cofactor of the L polymerase

P links L to the N-RNA template.

P associates with the L protein to form an active RNA-dependent RNA polymerase complex (Emerson and Yu 1975). In this complex, L carries out all enzymatic activities involved in the transcription and replication of the viral genome, including RNA synthesis, mRNA cap synthesis and methylation and mRNA polyadenylation, while P is an essential non-catalytic cofactor (Barr, Whelan et al. 2002). In the absence of P, L initiates RNA synthesis and produces short oligonucleotides of 2-4 ribonucleotides, but P is required for RNA elongation (De and Banerjee 1985). L is a processive enzyme, which must remain attached as it moves along its template (Stillman and Whitt 1999), but L alone has a weak binding affinity for the N-RNA complex and is unable to re-associate with its template (Mellon and Emerson 1978). P has binding sites for both L and the N-RNA complex and is presumed to correctly position L onto its N-RNA template and maintains it attached during its displacements along the template.

Binding site for L.

The binding site for the L protein is located in the N-terminal part of P, although its precise localization is unclear (Chenik, Schnell et al. 1998; Jacob, Real et al. 2001; Castel, Chteoui et al. 2009). The binding site was initially localized in the 60 first amino acids of P, which are required for transcription, whereas residues 61-173 were thought to be dispensable (Jacob, Real et al. 2001). Later, the binding site was suggested to be contained in the 19 N-terminal amino acids (Chenik, Schnell et al. 1998), but a recent study showed that the 100 N-terminal amino acids are required for a strong interaction with L and that the region 40-70 plays a major role (Castel, Chteoui et al. 2009). In VSV, a binding site for L was also localized in the distal part of the N-terminal IDR (aa 79-123) (Emerson and Schubert 1987). The requirement of a globally disordered acidic region for the correct assembly of the transcription/replication complex is reminiscent of eukaryotic transcription regulator proteins, such as GAL4 and GCN4. These proteins contain disordered acidic regions that activate transcription by recruiting transcription factors to the promoter and helping in the assembly of the transcription complex (Ptashne and Gann 1997). The acidic character of these regions, and not their specific amino acid sequence, is the major determinant of their activity, and transcriptional activation can be achieved with random polypeptides of identical composition (Ma and Ptashne 1987). In VSV, for transcription the N-terminal part of P can be replaced by β-tubulin, a protein with an acidic C-terminal region (Chattopadhyay and Banerjee 1988). It is also known that β-tubulin binds to the L polymerase (Moyer, Baker et al. 1986). The N-terminal disordered region of P (P_{NTD} + IDR_{NTD}) likely recruits the L polymerase to the site of RNA synthesis and allows the assembly of the multimolecular transcription and replication complexes.

Binding site for the N-RNA template.

The binding site for the nucleoprotein-RNA template was mapped in the C-terminal region of P (Chenik, Chebli et al. 1994; Fu, Zheng et al. 1994; Jacob, Real et al. 2001). This region is critical for RNA synthesis and corresponds closely to the last predicted structured region in RV P (Figure 4d). The crystal structures of P_{CTD} from RV and the related Mokola virus (*Lyssavirus*, genotype 3) revealed that this domain (aa 186-297 in RV) has the shape of a half-pear, the pear being cut lengthwise (Mavrakis, McCarthy et al. 2004; Assenberg, Delmas et al. 2009). Several residues involved in binding the N-RNA template, previously identified by a mutational analysis in Mokola virus, form patches on the opposite faces of the half-pear (Jacob, Real et al. 2001). Conserved hydrophobic residues line a hydrophobic cavity on the rounded face, named the W-hole, and four basic residues form a patch on the flat face. The corresponding P_{CTD} domain was correctly identified in VSV P from the meta-prediction analysis (aa 195-267), and its structure was solved by NMR spectroscopy (Ribeiro, Favier et al. 2008). A longer form of this domain (aa 185-267) confirmed that P_{CTD} is preceded by a flexible linker. Although multiple sequence alignments indicate almost no sequence similarity between P proteins from VSV and RV, the structure of VSV P_{CTD} is homologous to those of RV and Mokola Virus P and the structures can be aligned (Ribeiro, Favier et al. 2008).

Structure of the complex between P and the N-RNA template.

To understand how P attaches to the N-RNA template, we built a model of the complex formed by a recombinant circular N-RNA complex and P_{CTD} by computational modeling using SAXS and biochemical data as constraints (Ribeiro, Leyrat et al. 2009). In this model, P_{CTD} lies on the top of the C-terminal domain of one N protomer, while by an induced fit mechanism otherwise flexible loops from two adjacent N protomers (Albertini, Wernimont et al. 2006) mold around P_{CTD} and form extensive intermolecular contacts with the W-hole and the conserved basic patch, respectively. The involvement of two adjacent protomers indicates that the complete binding site for P_{CTD} only exists in multimeric N-RNA complexes. The binding of P_{CTD} to the N-RNA complex is non-cooperative with a dissociation constant near 100 nM (Ribeiro, Leyrat et al. 2009). The homologous complex of VSV solved by X-ray crystallography, showed a similar binding mode for P_{CTD} (Green and Luo 2009). The structure of both complexes revealed no conformational change in N that would significantly modify the accessibility to the RNA.

In RV, phosphorylation at S389 of N enhances the binding affinity for P_{CTD} (Toriumi and Kawai 2004). Our model suggests that phosphorylation of this serine residue creates an additional network of stabilizing electrostatic interactions (Ribeiro, Leyrat et al. 2009) and recent data show that phosphorylation of N achieves a 1000-fold increase in binding affinity (Ribeiro et al., unpublished data).

The overall structure of P and its roles in the transcription/replication complexes

Overall structure of RV P dimers.

P dimer is elongated with overall dimensions much larger than those expected for a globular protein of the same molecular mass (Gérard, Ribeiro et al. 2007). P has a dimeric central core from which four flexible "arms" emerge (Figure 5). So far no interaction involving the central domain has been described, but both the N-terminal and C-terminal flexible "arms" contain multiple binding sites. The N-terminal "arm" (P_{NTD} + IDR_{NTD}) contains no stable structured regions and behaves globally as a random coil chain. The C-terminal "arm" contains a terminal folded domain (P_{CTD}) tethered to P_{CED} by a long flexible linker (IDR_{CTD}). As in the N-terminal arm, the C-terminal flexible linker (IDR_{CTD}) contains a binding site for a cellular partner (see below). Conversely, the C-terminal domain is a well-structured domain, involved in the interaction with the N-RNA template but also containing binding sites for various cellular proteins (importin, exportin, kinase, STATs, PML and microtubules). The overlap of binding sites in both N- and C-terminal "arms" certainly modulates the capacity of RV to replicate and might be part of the strategy developed by the virus for limiting the impact of its replication on the cellular metabolism.

Role of P in RNA synthesis.

In the transcription/replication complexes, the high flexibility of P and the high binding

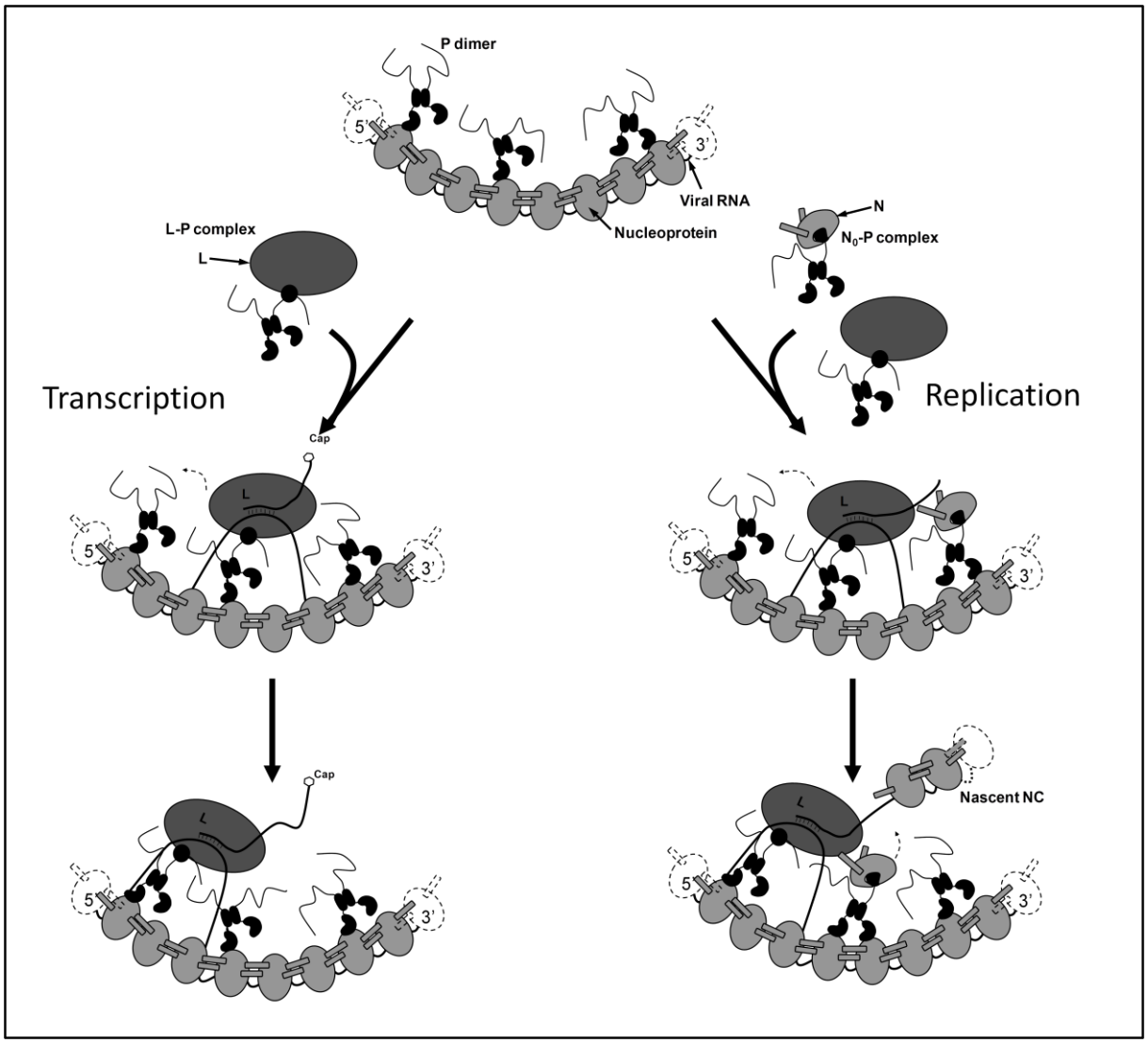


Figure 6. Schematic representation of the mechanisms of transcription and replication of RV. During RNA synthesis, the polymerase must move along its RNA template and must displace the RNA from the nucleoproteins. P provides the physical link by which the polymerase attaches to its template. In RV, the dimerization domain of P is dispensable for transcription and cannot therefore move by a cartwheeling mechanism. Because of the strong binding affinity of P_{CTD} for the N-RNA template, we suppose that P dimers are bound at regular intervals along the N-RNA complex. During transcription, the polymerase attaches to the N-terminal region of P and transcribes a stretch of RNA. Because P is elongated and flexible, the N-terminal arm of the adjacent P molecule can extend and catch the polymerase, which is then transferred downstream. The polymerase can thus move along the N-RNA complex by jumping between P molecules. During replication, the polymerase can move by a similar mechanism, but in addition, the newly synthesized genomic RNA is encapsidated by nascent N molecules. Nascent N molecules form soluble N^0-P complexes, which can also attach to the N-RNA complex through their P_{CTD}. N^0 can then be delivered to the site of RNA synthesis, yielding new N-RNA complexes.

affinity of P_{CTD} for the N-RNA template suggests an alternative mechanism for the movement of the polymerase along its N-RNA template, in which dimerization of P is dispensable (Figure 6). Highly flexible P molecules tethered to the N-RNA template through their P_{CTD} will sample large conformational spaces, acting like brushes that repel each other. If P molecules are distributed at regular intervals along the template, displaying their N^0 and L binding sites, the polymerase could move along the template by jumping from one P to the next. The long N-terminal arms of P could fetch the polymerase upstream, maintain it attached to the template during the time it replicates a stretch of RNA, and then deliver it downstream to the next P molecule. Similarly, the N-terminal arms could catch N^0 molecule and position them correctly for encapsidation of the newly synthesized RNA molecule.

Intracellular trafficking of P

Nucleo-cytoplasmic trafficking.

During the life cycle of viruses, viral components need to be transported in the cytoplasm or in and out the nucleus. Numerous viruses hijack

the cellular motors and the cytoskeleton tracks for intracellular transport as well as the cellular nucleo-cytoplasmic shuttles for proceeding through the nuclear pore complex (Leopold and Pfister 2006). Although rabies virus replicates in the cytoplasm, P shuttles between the cytoplasm and the nucleus through interactions with importins and exportins. The correct trafficking of P is critical for the interactions with host cells and for the pathogenicity of the virus (Ito, Moseley et al. 2010). The protein contains a bipartite nuclear localization sequence (NLS) (Dingwall and Laskey 1991), consisting in a conserved basic motif ($K_{211}KYK_{214}$) and Arg_{260}, that form a patch on the surface of the C-terminal domain (Pasdeloup, Poisson et al. 2005) (Figure 3). The protein also contains two conserved chromosome region maintenance protein 1 (CRM1)-dependent nuclear export signals (NES) (Bogerd, Fridell et al. 1996), one in the N-terminal region (aa 49-58) and the other in P_{CTD} (aa 224-234) (Pasdeloup, Poisson et al. 2005; Moseley, Filmer et al. 2007) (Figure 3). The N-terminal NES is localized in the disordered N-terminal part of P, whereas the C-terminal NES is close to the import sequence and to a phosphorylation site (S210). The residues forming the second NES are buried inside P_{CTD} structure, and the activation of this NES is controlled phosphorylation of the adjacent residue (Moseley, Filmer et al. 2007).

Interactions with the cytoskeleton.

Different studies revealed that RV and VSV P interact with the microtubules, but the precise role of these interactions remains poorly understood. RV P binds to the light chain 8 (LC8) of the host cell dynein complex (Jacob, Badrane et al. 2000; Raux, Flamand et al. 2000). Dynein is a multi-molecular complex that transports various cellular cargos towards the minus-end of cytoskeletal microtubules, usually oriented towards the cell center. This interaction thus suggests that RV uses the host cell machinery for intracellular transport (Jacob, Badrane et al. 2000; Raux, Flamand et al. 2000) and could possibly explain the retrograde axonal transport of RV. However, a recent study showed that the whole virus rather than the nucleocapsid is transported to the cell body, a process that cannot therefore depend on P, which is located inside the viral particle (Klingen, Conzelmann et al. 2008). Moreover, deletion of the LC8 binding region (aa 138-149) of P did not attenuate the spread of the virus from a peripheral entry site to the central nervous system (CNS) (Mebatsion 2001; Tan, Preuss et al. 2007). Dynein could also help the polymerase to move along the N-RNA template since deletion of the LC8 binding region (aa 138-149) reduces the production of viral RNA in the CNS (Mebatsion 2001; Tan, Preuss et al. 2007). However, mutation of two residues in the LC8-binding site P abolishes the interaction with LC8, but has no effect on transcription (Poisson, Real et al. 2001). More recently, it was shown that the association of P3 with microtubules participates in its NLS-dependent nuclear import (Moseley et al., 2007b). In VSV, no binding site for LC8 has been found, but P protein also interacts with microtubules and participates in the transport of nucleocapsids towards the cell periphery *via* a microtubule-mediated process (Das et al., 2006).

P is a multifunctional antagonist of the host cell interferon response

The cellular interferon response.

The innate immune response represents the first line of defense of an organism against viral infection. The recognition of viral components by Toll-like receptors (TLRs) or retinoic acid-inducible-like receptors (RLRs) triggers the production of interferons. Interferons activate signaling cascades and induce the expression of various interferon-stimulated genes (ISGs) with antiviral properties. Many viruses, including RV, have developed strategies for evading these cellular antiviral responses, in particular the production of type I interferon (Chelbi-Alix, Vidy et al. 2006; Katze, Fornek et al. 2008; Rieder and Conzelmann 2009). RV induces an innate immune response (Wang, Sarmento et al. 2005; Hornung, Ellegast et al. 2006) and is sensitive to interferon treatment (Brzozka, Finke et al. 2005; Delhaye, Paul et al. 2006). P plays a major role in modulating this response (Finke, Brzozka et al. 2004; Shimizu, Ito et al. 2006). An engineered RV strain producing insufficient amounts of P to inhibit the host-cell response could replicate in interferon-deficient cells but not in interferon-competent cells (Brzozka, Finke et al. 2005; Brzozka, Finke et al. 2006).

P blocks the interferon response.

P acts at different stages of the interferon pathway. Firstly, P interferes with the production of interferon by preventing the phosphorylation of the

interferon regulatory factor 3 (IRF-3) by the TANK-binding kinase 1 (Brzozka, Finke et al. 2005). Secondly, P interferes with the interferon-stimulated JAK/STAT signaling pathway by blocking the nuclear import of the activated signal transducer and activator of transcription 1 and 2 (STAT1 and STAT2) (Vidy, Chelbi-Alix et al. 2005; Brzozka, Finke et al. 2006). P acts downstream of the activation of STATs. It prevents their nuclear import by sequestering phosphorylated STATs in the cytoplasm through interactions with the microtubules (Moseley, Lahaye et al. 2009) and it prevents STATs from binding to DNA in the nucleus (Vidy, Bougrini et al. 2007). Impairing the ability of P to retain STATs in the cytoplasm reduces the pathogenecity of the RV strain probably by allowing more efficient interferon response (Ito, Moseley et al. 2010). P_{CTD}, in particular the last ten amino acids, is required for the interaction with STATs (Vidy, Chelbi-Alix et al. 2005; Brzozka, Finke et al. 2006) (Figure 3), but not for inhibiting the transcriptional activation of IRF-3, indicating that another region of P is involved in the interaction with this other transcription factor. Thirdly, P interferes with ISG products such as the pro-myelocytic leukemia (PML) protein (also known as the tripartite motif protein 19, TRIM19). The PML protein is part of nuclear multi-protein complexes, named nuclear bodies that are involved in host cell defense mechanisms (Bernardi and Pandolfi 2007). Although the mechanism of antiviral action of these cellular components remains unknown, nuclear bodies are involved in the interferon-stimulated host-cell response against RV infection. In RV infected cells, nuclear bodies grow bigger, and in the absence of PML expression, higher titers of RV are obtained suggesting that the absence of a particular PML isoform (isoform IV) enhances virus replication (Blondel, Regad et al. 2002; Blondel, Kheddache et al. 2010). The expression of the truncated P3 induces modifications of the nuclear bodies similar to those observed in infected cells, while expression of the entire P protein retains PML in the cytoplasm (Blondel, Regad et al. 2002). The interaction with RV P occurs between P_{CTD} and the "really interesting new gene" (RING) finger domain of the PML protein (Figure 3).

Future perspective

Recent studies of RV P revealed the peculiar structural properties of this protein. The existence of two folded domains and two long intrinsically disordered regions confers the possibility of continuous extension and contraction of the polypeptide chain. The presence of binding sites for multiple viral and cellular partners, including in the IDRs, backs the role of P as an adapter in the formation of several multi-molecular complexes. This new picture provides a first glimpse at the way P plays its role in the viral transcription/replication machineries. However, numerous questions about the molecular mechanisms of P remain unanswered and it is necessary now to dig more deeply into the mechanisms of action of P. This will require combining more structural and biophysical studies of isolated proteins and protein complexes with *in vitro* reconstitutions of functional machineries and *in vivo* observation of these machineries in their physiological environment.

The structure of the complex between P_{CTD} and a circular N-RNA complex showed how P attaches to the transcription/replication template. We must now determine if P moves along the N-RNA template, serving as a carrier of L or if it remains static and forms a chain where L is transmitted from one to the other. In both cases, the L-P polymerase moves along the template, and we must determine what drives P or L. Resolving the structure of the N-RNA-PCTD complex showed that binding of P_{CTD} to the N-RNA template does not modify the accessibility of the RNA. We need now to understand how the polymerase gets access to the RNA. We found that the N-terminal N^0 binding region of P is globally disordered but possesses some residual structure that could fold upon binding to nascent N molecules. We need to find out how P binding to N^0, prevents the binding of cellular RNAs and allows the specific encapsidation of viral RNAs. We must also understand how N is transferred from P to the viral RNA and how the new nucleocapsids assemble. In the host cell, transcription and replication of the viral genome occur in Negri bodies. We need to characterize the organization of the transcription/replication machineries within these cellular compartments and their interactions with cellular components and understand how viral transcription and replication occur in the host cell. Recent studies disclosed important interactions between RV P and host cell components. We need to delve more deeply into the molecular mechanisms of these interactions to understand

how RV precisely diverts these cellular machineries. Finally, deciphering the molecular mechanisms by which P interacts with viral and cellular partners should also uncover new targets for the development of an antiviral treatment.

References

Albertini, A. A., A. K. Wernimont, et al. (2006). "Crystal structure of the rabies virus nucleoprotein-RNA complex." Science **313**(5785): 360-3.

Assenberg, R., O. Delmas, et al. (2009). "The Structure of the N-RNA Binding Domain of the Mokola virus Phosphoprotein." J Virol: Nov 11. [Epub ahead of print]PMID: 19906936 [PubMed - as supplied by publisher].

Barik, S. and A. K. Banerjee (1992). "Phosphorylation by cellular casein kinase II is essential for transcriptional activity of vesicular stomatitis virus phosphoprotein P." Proc Natl Acad Sci U S A **89**(14): 6570-4.

Barik, S. and A. K. Banerjee (1992). "Sequential phosphorylation of the phosphoprotein of vesicular stomatitis virus by cellular and viral protein kinases is essential for transcription activation." J Virol **66**(2): 1109-18.

Barr, J. N., S. P. Whelan, et al. (2002). "Transcriptional control of the RNA-dependent RNA polymerase of vesicular stomatitis virus." Biochim Biophys Acta **1577**(2): 337-53.

Bernardi, R. and P. P. Pandolfi (2007). "Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies." Nat Rev Mol Cell Biol **8**(12): 1006-16.

Blondel, D., S. Kheddache, et al. (2010). "Resistance to Rabies viral infection conferred by PMLIV isoform." J Virol.

Blondel, D., T. Regad, et al. (2002). "Rabies virus P and small P products interact directly with PML and reorganize PML nuclear bodies." Oncogene **21**(52): 7957-70.

Bogerd, H. P., R. A. Fridell, et al. (1996). "Protein sequence requirements for function of the human T-cell leukemia virus type 1 Rex nuclear export signal delineated by a novel in vivo randomization-selection assay." Mol Cell Biol **16**(8): 4207-14.

Brzozka, K., S. Finke, et al. (2005). "Identification of the rabies virus alpha/beta interferon antagonist: phosphoprotein P interferes with phosphorylation of interferon regulatory factor 3." J Virol **79**(12): 7673-81.

Brzozka, K., S. Finke, et al. (2006). "Inhibition of interferon signaling by rabies virus phosphoprotein P: activation-dependent binding of STAT1 and STAT2." J Virol **80**(6): 2675-83.

Castel, G., M. Chteoui, et al. (2009). "Peptides that mimic the amino-terminal end of the rabies virus phosphoprotein have antiviral activity." J Virol **83**(20): 10808-20.

Chattopadhyay, D. and A. K. Banerjee (1987). "Phosphorylation within a specific domain of the phosphoprotein of vesicular stomatitis virus regulates transcription in vitro." Cell **49**(3): 407-14.

Chattopadhyay, D. and A. K. Banerjee (1988). "NH2-terminal acidic region of the phosphoprotein of vesicular stomatitis virus can be functionally replaced by tubulin." Proc Natl Acad Sci U S A **85**(21): 7977-81.

Chelbi-Alix, M. K., A. Vidy, et al. (2006). "Rabies viral mechanisms to escape the IFN system: the viral protein P interferes with IRF-3, Stat1, and PML nuclear bodies." J Interferon Cytokine Res **26**(5): 271-80.

Chen, M., T. Ogino, et al. (2007). "Interaction of vesicular stomatitis virus P and N proteins: Identification of two overlapping domains at the N-terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA." J Virol **81**: 13478-13485.

Chenik, M., K. Chebli, et al. (1995). "Translation initiation at alternate in-frame AUG codons in the rabies virus phosphoprotein mRNA is mediated by a ribosomal leaky scanning mechanism." J Virol **69**(2): 707-12.

Chenik, M., K. Chebli, et al. (1994). "In vivo interaction of rabies virus phosphoprotein (P) and nucleoprotein (N): existence of two N-binding sites on P protein." J Gen Virol **75** 2889-96.

Chenik, M., M. Schnell, et al. (1998). "Mapping the interacting domains between the rabies virus polymerase and phosphoprotein." J Virol **72**(3): 1925-30.

Childs, K. S., J. Andrejeva, et al. (2009). "Mechanism of mda-5 Inhibition by paramyxovirus V proteins." J Virol **83**(3): 1465-73.

Cleaveland, S., E. M. Fevre, et al. (2002). "Estimating human rabies mortality in the United Republic of Tanzania from dog bite injuries." Bull World Health Organ **80**(4): 304-10.

Dancho, B., M. O. McKenzie, et al. (2009). "Vesicular stomatitis virus matrix protein mutations that affect association with host membranes and viral nucleocapsids." J Biol Chem **284**(7): 4500-9.

Das, S. C. and A. K. Pattnaik (2005). "Role of the hypervariable hinge region of phosphoprotein P of vesicular stomatitis virus in viral RNA synthesis and assembly of infectious virus particles." J Virol **79**(13): 8101-12.

Das, T., A. K. Gupta, et al. (1995). "Role of cellular casein kinase II in the function of the phosphoprotein (P) subunit of RNA polymerase of vesicular stomatitis virus." J Biol Chem **270**(41): 24100-7.

Das, T., A. K. Pattnaik, et al. (1997). "Basic amino acid residues at the carboxy-terminal eleven amino acid region of the phosphoprotein (P) are required for transcription but not for replication of vesicular stomatitis virus genome RNA." Virology **238**(1): 103-14.

De, B. P. and A. K. Banerjee (1985). "Requirements and functions of vesicular stomatitis virus L and NS proteins in the transcription process in vitro." Biochem Biophys Res Commun **126**(1): 40-9.

Delhaye, S., S. Paul, et al. (2006). "Neurons produce type I interferon during viral encephalitis." Proc Natl Acad Sci U S A **103**(20): 7835-40.

Dietzschold, B., J. Li, et al. (2008). "Concepts in the pathogenesis of rabies." Future Virol **3**(5): 481-490.

Ding, H., T. J. Green, et al. (2006). "Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus." J Virol **80**(6): 2808-14.

Ding, H., T. J. Green, et al. (2004). "Crystallization and preliminary X-ray analysis of a proteinase-K-resistant domain within the phosphoprotein of vesicular stomatitis virus (Indiana)." Acta Crystallogr D Biol Crystallogr **60**(Pt 11): 2087-90.

Dingwall, C. and R. A. Laskey (1991). "Nuclear targeting sequences--a consensus?" Trends Biochem Sci **16**(12): 478-81.

Dyson, H. J. and P. E. Wright (2002). "Coupling of folding and binding for unstructured proteins." Curr Opin Struct Biol **12**(1): 54-60.

Emerson, S. U. and M. Schubert (1987). "Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus." Proc Natl Acad Sci U S A **84**(16): 5655-9.

Emerson, S. U. and Y. Yu (1975). "Both NS and L proteins are required for in vitro RNA synthesis by vesicular stomatitis virus." J Virol **15**(6): 1348-56.

Finke, S., K. Brzozka, et al. (2004). "Tracking fluorescence-labeled rabies virus: enhanced green fluorescent protein-tagged phosphoprotein P supports virus gene expression and formation of infectious particles." J Virol **78**(22): 12333-43.

Finke, S. and K. K. Conzelmann (2003). "Dissociation of rabies virus matrix protein functions in regulation of viral

RNA synthesis and virus assembly." J Virol **77**(22): 12074-82.
Fooks, A. R. (2005). "Rabies remains a 'neglected disease'." Euro Surveill **10**(11): 211-2.
Fu, Z. F., Y. Zheng, et al. (1994). "Both the N- and the C-terminal domains of the nominal phosphoprotein of rabies virus are involved in binding to the nucleoprotein." Virology **200**(2): 590-7.
Gao, Y. and J. Lenard (1995). "Cooperative binding of multimeric phosphoprotein (P) of vesicular stomatitis virus to polymerase (L) and template: pathways of assembly." J Virol **69**(12): 7718-23.
Gao, Y. and J. Lenard (1995). "Multimerization and transcriptional activation of the phosphoprotein (P) of vesicular stomatitis virus by casein kinase-II." Embo J **14**(6): 1240-7.
Gaudin, Y., C. Tuffereau, et al. (1999). "Rabies virus-induced membrane fusion." Mol Membr Biol **16**(1): 21-31.
Ge, P., J. Tsao, et al. (2010). "Cryo-EM model of the bullet-shaped vesicular stomatitis virus." Science **327**(5966): 689-93.
Gérard, F., E. Ribeiro, et al. (2007). "Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution." Biochemistry **46**: 10328-10338.
Gérard, F. C. A., E. A. Ribeiro , et al. (2009). "Modular organization of rabies virus phosphoprotein." J. Mol. Biol. **388**: 978-996.
Gigant, B., F. Iseni, et al. (2000). "Neither phosphorylation nor the amino-terminal part of rabies virus phosphoprotein is required for its oligomerization." J Gen Virol **81**(Pt 7): 1757-61.
Gill, D. S., D. Chattopadhyay, et al. (1986). "Identification of a domain within the phosphoprotein of vesicular stomatitis virus that is essential for transcription in vitro." Proc Natl Acad Sci U S A **83**(23): 8873-7.
Green, T. J. and M. Luo (2009). "Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P." Proc Natl Acad Sci U S A **106**: 11721-11726.
Green, T. J., X. Zhang, et al. (2006). "Structure of the vesicular stomatitis virus nucleoprotein-RNA complex." Science **313**(5785): 357-60.
Gupta, A. K., D. Blondel, et al. (2000). "The phosphoprotein of rabies virus is phosphorylated by a unique cellular protein kinase and specific isomers of protein kinase C." J Virol **74**(1): 91-8.
Hanlon, C. A., I. V. Kuzmin, et al. (2005). "Efficacy of rabies biologics against new lyssaviruses from Eurasia." Virus Res **111**(1): 44-54.
Harty, R. N., J. Paragas, et al. (1999). "A proline-rich motif within the matrix protein of vesicular stomatitis virus and rabies virus interacts with WW domains of cellular proteins: implications for viral budding." J Virol **73**(4): 2921-9.
He, B., K. Wang, et al. (2009). "Predicting intrinsic disorder in proteins: an overview." Cell Res **19**(8): 929-49.
Hemachudha, T., J. Laothamatas, et al. (2002). "Human rabies: a disease of complex neuropathogenetic mechanisms and diagnostic challenges." Lancet Neurol **1**(2): 101-9.
Hornung, V., J. Ellegast, et al. (2006). "5'-Triphosphate RNA is the ligand for RIG-I." Science **314**(5801): 994-7.
Howard, M. and G. Wertz (1989). "Vesicular stomatitis virus RNA replication: a role for the NS protein." J Gen Virol **70 (Pt 10)**: 2683-94.
Hudson, L. D., C. Condra, et al. (1986). "Cloning and expression of a viral phosphoprotein: structure suggests vesicular stomatitis virus NS may function by mimicking an RNA template." J Gen Virol **67 (Pt 8)**: 1571-9.
Hwang, L. N., N. Englund, et al. (1999). "Optimal replication activity of vesicular stomatitis virus RNA polymerase requires phosphorylation of a residue(s) at carboxy-terminal domain II of its accessory subunit, phosphoprotein P." J Virol **73**(7): 5613-20.
Iseni, F., F. Baudin, et al. (2000). "Structure of the RNA inside the vesicular stomatitis virus nucleocapsid." Rna **6**(2): 270-81.
Ito, N., G. W. Moseley, et al. (2010). "Role of interferon antagonist activity of rabies virus phosphoprotein in viral pathogenicity." J Virol **84**(13): 6699-710.
Ivanov, I., T. Crepin, et al. (2010). "Structure of the dimerisation domain of the rabies virus phosphoprotein." J. Virol. **84**: 3707-3710.
Jacob, Y., H. Badrane, et al. (2000). "Cytoplasmic dynein LC8 interacts with lyssavirus phosphoprotein." J Virol **74**(21): 10217-22.
Jacob, Y., E. Real, et al. (2001). "Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains." J Virol **75**(20): 9613-22.
Jensen, M. R., K. Houben, et al. (2008). "Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein." J Am Chem Soc **130**(25): 8055-61.
Karlin, D., F. Ferron, et al. (2003). "Structural disorder and modular organization in Paramyxovirinae N and P." J Gen Virol **84**(Pt 12): 3239-52.
Katze, M. G., J. L. Fornek, et al. (2008). "Innate immune modulation by RNA viruses: emerging insights from functional genomics." Nat Rev Immunol **8**(8): 644-54.
Klingen, Y., K. K. Conzelmann, et al. (2008). "Double-labeled rabies virus: live tracking of enveloped virus transport." J Virol **82**(1): 237-45.
Knobel, D. L., S. Cleaveland, et al. (2005). "Re-evaluating the burden of rabies in Africa and Asia." Bull World Health Organ **83**(5): 360-8.
Kolakofsky, D., P. Le Mercier, et al. (2004). "Viral RNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis." Virology **318**(2): 463-73.
Komarova, A. V., E. Real, et al. (2007). "Rabies virus matrix protein interplay with eIF3, new insights into rabies virus pathogenesis." Nucleic Acids Res **35**(5): 1522-32.
Lahaye, X., A. Vidy, et al. (2009). "Functional characterization of Negri bodies (NBs) in rabies virus-infected cells: Evidence that NBs are sites of viral transcription and replication." J Virol **83**(16): 7948-58.
Leopold, P. L. and K. K. Pfister (2006). "Viral strategies for intracellular trafficking: motors and microtubules." Traffic **7**(5): 516-23.
Leslie, M. J., S. Messenger, et al. (2006). "Bat-associated rabies virus in Skunks." Emerg Infect Dis **12**(8): 1274-7.
Lichty, B. D., A. T. Power, et al. (2004). "Vesicular stomatitis virus: re-inventing the bullet." Trends Mol Med **10**(5): 210-6.
Ma, J. and M. Ptashne (1987). "Deletion analysis of GAL4 defines two transcriptional activating segments." Cell **48**(5): 847-53.
Mavrakis, M., F. Iseni, et al. (2003). "Isolation and characterisation of the rabies virus N degrees-P complex produced in insect cells." Virology **305**(2): 406-14.
Mavrakis, M., A. A. McCarthy, et al. (2004). "Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus." J Mol Biol **343**(4): 819-31.
Mavrakis, M., S. Mehouas, et al. (2006). "Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA." Virology **349**(2): 422-9.
Mebatsion, T. (2001). "Extensive attenuation of rabies virus by simultaneously modifying the dynein light chain binding site in the P protein and replacing Arg333 in the G protein." J Virol **75**(23): 11496-502.
Mellon, M. G. and S. U. Emerson (1978). "Rebinding of transcriptase components (L and NS proteins) to the nucleocapsid template of vesicular stomatitis virus." J Virol **27**(3): 560-7.

Mohan, A., C. J. Oldfield, et al. (2006). "Analysis of molecular recognition features (MoRFs)." J Mol Biol **362**(5): 1043-59.

Moseley, G. W., R. P. Filmer, et al. (2007). "Nucleocytoplasmic distribution of rabies virus P-protein is regulated by phosphorylation adjacent to C-terminal nuclear import and export signals." Biochemistry **46**(43): 12053-61.

Moseley, G. W., X. Lahaye, et al. (2009). "Dual modes of rabies P-protein association with microtubules: a novel strategy to suppress the antiviral response." J Cell Sci **122**(Pt 20): 3652-62.

Moseley, G. W., D. M. Roth, et al. (2007). "Dynein light chain association sequences can facilitate nuclear protein import." Mol Biol Cell **18**(8): 3204-13.

Moyer, S. A., S. C. Baker, et al. (1986). "Tubulin: a factor necessary for the synthesis of both Sendai virus and vesicular stomatitis virus RNAs." Proc Natl Acad Sci U S A **83**(15): 5405-9.

Ogino, T. and A. K. Banerjee (2007). "Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus." Mol Cell **25**(1): 85-97.

Pasdeloup, D., N. Poisson, et al. (2005). "Nucleocytoplasmic shuttling of the rabies virus P protein requires a nuclear localization signal and a CRM1-dependent nuclear export signal." Virology **334**(2): 284-93.

Pattnaik, A. K., L. Hwang, et al. (1997). "Phosphorylation within the amino-terminal acidic domain I of the phosphoprotein of vesicular stomatitis virus is required for transcription but not for replication." J Virol **71**(11): 8167-75.

Patton, J. T., N. L. Davis, et al. (1984). "N protein alone satisfies the requirement for protein synthesis during RNA replication of vesicular stomatitis virus." J Virol **49**(2): 303-9.

Paul, P. R., D. Chattopadhyay, et al. (1988). "The functional domains of the phosphoprotein (NS) of vesicular stomatitis virus (Indiana serotype)." Virology **166**(2): 350-7.

Pei, J. and N. V. Grishin (2001). "AL2CO: calculation of positional conservation in a protein sequence alignment." Bioinformatics **17**(8): 700-12.

Peluso, R. W. and S. A. Moyer (1988). "Viral proteins required for the in vitro replication of vesicular stomatitis virus defective interfering particle genome RNA." Virology **162**(2): 369-76.

Poisson, N., E. Real, et al. (2001). "Molecular basis for the interaction between rabies virus phosphoprotein P and the dynein light chain LC8: dissociation of dynein-binding properties and transcriptional functionality of P." J Gen Virol **82**(Pt 11): 2691-6.

Pringle, C. R. (1997). "The order Mononegavirales--current status." Arch Virol **142**(11): 2321-6.

Ptashne, M. and A. Gann (1997). "Transcriptional activation by recruitment." Nature **386**(6625): 569-77.

Qanungo, K. R., D. Shaji, et al. (2004). "Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA." Proc Natl Acad Sci U S A **101**(16): 5952-7.

Raux, H., A. Flamand, et al. (2000). "Interaction of the rabies virus P protein with the LC8 dynein light chain." J Virol **74**(21): 10212-6.

Ribeiro, E. A., Jr., A. Favier, et al. (2008). "Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein." J Mol Biol **382**(2): 525-38.

Ribeiro, E. A., Jr., A. Favier, et al. (2008). "Solution Structure of the C-Terminal Nucleoprotein-RNA Binding Domain of the Vesicular Stomatitis Virus Phosphoprotein." J Mol Biol **382**: 525-538.

Ribeiro, E. A., C. Leyrat, et al. (2009). "Binding of rabies virus polymerase cofactor to recombinant circular nucleoprotein-RNA complexes." J Mol Biol **394**(3): 558-75.

Rieder, M. and K. K. Conzelmann (2009). "Rhabdovirus evasion of the interferon system." J Interferon Cytokine Res **29**(9): 499-509.

Roche, S., S. Bressanelli, et al. (2006). "Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G." Science **313**(5784): 187-91.

Roche, S., F. A. Rey, et al. (2007). "Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G." Science **315**(5813): 843-8.

Rose, N. F., J. Publicover, et al. (2008). "Hybrid alphavirus-rhabdovirus propagating replicon particles are versatile and potent vaccine vectors." Proc Natl Acad Sci U S A **105**(15): 5839-43.

Rupprecht, C. E., R. Willoughby, et al. (2006). "Current and future trends in the prevention, treatment and control of rabies." Expert Rev Anti Infect Ther **4**(6): 1021-38.

Schnell, M. J., J. P. McGettigan, et al. (2010). "The cell biology of rabies virus: using stealth to reach the brain." Nat Rev Microbiol **8**(1): 51-61.

Schoehn, G., F. Iseni, et al. (2001). "Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site." J Virol **75**(1): 490-8.

Shimizu, K., N. Ito, et al. (2006). "Sensitivity of rabies virus to type I interferon is determined by the phosphoprotein gene." Microbiol Immunol **50**(12): 975-8.

Spolar, R. S. and M. T. Record, Jr. (1994). "Coupling of local folding to site-specific binding of proteins to DNA." Science **263**(5148): 777-84.

Stillman, E. A. and M. Whitt (1999). "Transcript initiation and 5'-end modifications are separable events during vesicular stomatitis virus transcription." J. Virol. **73**: 7199-7209.

Streicker, D. G., A. S. Turmelle, et al. (2010). "Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats." Science **329**(5992): 676-9.

Takacs, A. M., T. Das, et al. (1993). "Mapping of interacting domains between the nucleocapsid protein and the phosphoprotein of vesicular stomatitis virus by using a two-hybrid system." Proc Natl Acad Sci U S A **90**(21): 10375-9.

Tan, G. S., M. A. Preuss, et al. (2007). "The dynein light chain 8 binding motif of rabies virus phosphoprotein promotes efficient viral transcription." Proc Natl Acad Sci U S A **104**(17): 7229-34.

Théodoridès, J. (1986). Histoire de la rage. Paris, Masson.

Tokuriki, N., C. J. Oldfield, et al. (2009). "Do viral proteins possess unique biophysical features?" Trends Biochem Sci **34**(2): 53-9.

Tompa, P. (2002). "Intrinsically unstructured proteins." Trends Biochem Sci **27**(10): 527-33.

Toriumi, H. and A. Kawai (2004). "Association of rabies virus nominal phosphoprotein (P) with viral nucleocapsid (NC) is enhanced by phosphorylation of the viral nucleoprotein (N)." Microbiol Immunol **48**(5): 399-409.

Uversky, V. N., C. J. Oldfield, et al. (2005). "Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling." J Mol Recognit **18**(5): 343-84.

Vidy, A., J. Bougrini, et al. (2007). "The nulceocytoplasmic rabies P protein counteracts interferons signaling by inhibiting both nulear accumulation and DNA binding of STAT1." J Virol. **81**: 4255-4263.

Vidy, A., M. Chelbi-Alix, et al. (2005). "Rabies virus P protein interacts with STAT1 and inhibits interferon signal transduction pathways." J Virol **79**(22): 14411-20.

Vidy, A., J. El Bougrini, et al. (2007). "The nucleocytoplasmic rabies virus P protein counteracts interferon signaling by inhibiting both nuclear accumulation and DNA binding of STAT1." J Virol **81**(8): 4255-63.

Wang, Z. W., L. Sarmento, et al. (2005). "Attenuated rabies virus activates, while pathogenic rabies virus evades, the host innate immune responses in the central nervous system." J Virol **79**(19): 12554-65.

Weyer, J., I. V. Kuzmin, et al. (2008). "Cross-protective and cross-reactive immune responses to recombinant vaccinia viruses expressing full-length lyssavirus glycoprotein genes." Epidemiol Infect **136**(5): 670-8.

Wirblich, C., G. S. Tan, et al. (2008). "PPEY motif within the rabies virus (RV) matrix protein is essential for efficient virion release and RV pathogenicity." J Virol **82**(19): 9730-8.

Wright, P. E. and H. J. Dyson (1999). "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm." J Mol Biol **293**(2): 321-31.

Wright, P. E. and H. J. Dyson (2009). "Linking folding and binding." Curr Opin Struct Biol **19**(1): 31-8.

Executive summary

Rabies and the rabies virus

- Rabies has been feared by mankind for more than 20 centuries and it remains an important but neglected infectious disease. The encephalitis remains incurable but is preventable by pre- and post-infection treatments.
- The main causative agent is the rabies virus, a non-segmented negative sense RNA virus.

The rabies virus replication complex

- Transcription and replication of RV occurs in the cytoplasm.
- The RNA genome and three viral proteins, N, P and L, constitute the infectious part of the virus and are required for an efficient and regulated RNA synthesis.
- The N-RNA complex serves as a template for both processes.
- L carries out all enzymatic activities, including RNA synthesis, mRNA capping, methylation and poly-adenylation.
- P is an essential non-catalytic cofactor for both transcription and replication.

Phosphoprotein functions

- P acts as a chaperone of nascent N molecules by forming a N^0-P complex that prevents N to bind non-specifically to cellular RNA and maintains it in a soluble form until it is used to encapsidate newly synthesized RNA genomes or antigenomes.
- P attaches the polymerase complexes to the N-RNA template and is required for the elongation of RNA molecules.
- P shuttles between the cytoplasm and the nucleus, hijacking cellular importins and exportins.
- P binds to the molecular motor dynein through direct interaction with the light chain LC8.
- P blocks the host cell innate immune response by interacting at different stages of the interferon pathway.

Phosphoprotein structure

- P forms elongated dimers.
- P contains two folded domains, a central dimerization domain and a C-terminal N-RNA binding domain, and long intrinsically disordered regions.
- Binding sites for several viral and cellular partners are located both in the disordered and folded regions.
- Folding upon binding to a partner could confer a high specificity with a moderate binding affinity, which might be important for the dynamic functioning of the transcription/replication complexes.

III. Objectifs du travail de thèse

L'objectif majeur du travail réalisé pendant cette thèse était de poursuivre la caractérisation biophysique, bioinformatique et structurale de la phosphoprotéine des *Rhabdoviridae* et de ses interactions avec la nucléoprotéine (notamment chez VSV et RV). En effet, la phosphoprotéine P constitue un composant majeur du complexe de réplication, peu caractérisé et impliqué dans la formation de multiples complexes avec la nucléoprotéine N^0, la matrice N-ARN, et la polymérase virale L. L'étude structurale de la protéine P, déjà entreprise à mon arrivée au laboratoire par Aurélie Albertini puis Francine Gérard et Euripedes de Almeida Ribeiro Jr, avait permis de déterminer que les phosphoprotéines des *Rhabdoviridae* existaient sous la forme de dimères allongés en solution (Gérard, Ribeiro et al. 2007). Très vite, la découverte de la nature modulaire de la phosphoprotéine nous a conduit à concentrer nos efforts sur l'étude de régions fonctionnelles isolées de la phosphoprotéine et à caractériser les interactions entre les modules N- et C-terminaux et la protéine N. Ce travail peut être subdivisé en plusieurs chapitres :

Les travaux sur l'organisation modulaire de la phosphoprotéine **(Chap. II)** définissent les différentes régions structurales et fonctionnelles de la protéine en termes de structure primaire, et ouvrent la voie vers une caractérisation plus poussée de la région N-terminale **(Chap. III)**, peu étudiée jusqu'à présent. La caractérisation structurale des domaines isolés de la protéine P (Mavrakis, McCarthy et al. 2004; Ding, Green et al. 2006; Ribeiro, Favier et al. 2008; Ivanov, Crepin et al. 2010), et notamment du domaine C-terminal de la phosphoprotéine du VSV **(Chap. IV)**, combinée à la connaissance des régions désordonnées, a permis la modélisation de la structure complète en solution sur base des données de diffusion de rayons X aux petits angles **(Chap. V)**. La connaissance de la structure de la protéine a ouvert quant à elle différentes perspectives pour l'étude des interactions du module fonctionnel C-terminal avec le complexe nucléoprotéine – ARN **(Chap. VI)**, et du module fonctionnel N-terminal avec la nucléoprotéine soluble N^0 **(Chap. VII)**, respectivement.

CHAPITRE II : L'ORGANISATION MODULAIRE DE LA PHOSPHOPROTEINE

gcq#122: "You Don't Wanna Know"

(Pulp Fiction)

I. Introduction

Les résultats exposés dans cet article définissent le contexte des travaux effectués au cours de cette thèse. En effet, l'organisation modulaire de la phosphoprotéine prédite par une méta-prédiction de régions désordonnées a été validée expérimentalement par le clonage, l'expression, la purification et la caractérisation biophysique des domaines structurés.

L'étude révèle la présence d'un domaine de dimérisation replié dans la région centrale de la molécule, ainsi que l'existence d'un autre domaine replié autonome dans la région C-terminale, impliqué dans la fixation au complexe N-ARN. Les frontières prédites entre régions désordonnées et domaines structurés sont en accord avec les structures cristallographiques disponibles pour le domaine central P_{CED} chez le VSV (Ding, Green et al. 2006), le domaine C-terminal P_{CTD} chez RV (Mavrakis, McCarthy et al. 2004), et les domaines P_{CED} et P_{CTD} du virus Sendaï (SV) (Tarbouriech, Curran et al. 2000; Blanchard, Tarbouriech et al. 2004). Ces travaux permettent également de définir précisément les frontières des domaines dont la structure est inconnue tels que RV P_{CED} et VSV P_{CTD}. La région N-terminale conservée de la phosphoprotéine (P_{NTD}), capable d'interagir avec la polymérase L (Castel, Chteoui et al. 2009) et la nucléoprotéine libre N^0 (Mavrakis, Mehouas et al. 2006; Chen, Ogino et al. 2007), possède d'après les analyses bioinformatiques une propension à former des hélices α. Cependant, un peptide correspondant aux résidus 1 à 68 de la phosphoprotéine du Virus de la Rage se comporte comme une protéine désordonnée en solution avec un contenu en hélice α de l'ordre de 5%.

L'organisation modulaire de la phosphoprotéine des *Rhabdoviridae* est conservée et possède d'importantes similarités avec les phosphoprotéines des *Paramyxoviridae* (Karlin, Ferron et al. 2003; Habchi, Mamelli et al. 2010).

II. ARTICLE I: Modular Organization of Rabies Virus Phosphoprotein

Auteurs: Francine Gérard, Euripedes de Almeida Ribeiro Jr, Cédric Leyrat, Ivan Ivanov, Danielle Blondel, Sonia Longhi, Rob Ruigrok, et Marc Jamin.

Publié en 2009 dans *Journal of Molecular Biology*, Volume 388, pages 978-996

doi:10.1016/j.jmb.2009.03.061 J. Mol. Biol. (2009) 388, 978–996

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Modular Organization of Rabies Virus Phosphoprotein

Francine C. A. Gerard[1]†, Euripedes de Almeida Ribeiro Jr[1]†, Cédric Leyrat[1], Ivan Ivanov[1], Danielle Blondel[2], Sonia Longhi[3], Rob W. H. Ruigrok[1] and Marc Jamin[1]*

[1]*UJF-EMBL-CNRS UMI 3265 - Unit of Virus Host Cell Interactions, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France*

[2]*Laboratoire de Virologie Moléculaire et Structurale, UMR 2472 CNRS-1157 INRA Bât. 14B, 1 Avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France*

[3]*Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS et Universités Aix-Marseille I et II, Case 932, 163 Avenue de Luminy, 13288 Marseille Cedex 09, France*

Received 12 November 2008; received in revised form 23 March 2009; accepted 25 March 2009
Available online 31 March 2009

Edited by J. Karn

A phosphoprotein (P) is found in all viruses of the *Mononegavirales* order. These proteins form homo-oligomers, fulfil similar roles in the replication cycles of the various viruses, but differ in their length and oligomerization state. Sequence alignments reveal no sequence similarity among proteins from viruses belonging to the same family. Sequence analysis and experimental data show that phosphoproteins from viruses of the *Paramyxoviridae* contain structured domains alternating with intrinsically disordered regions. Here, we used predictions of disorder of secondary structure, and an analysis of sequence conservation to predict the domain organization of the phosphoprotein from Sendai virus, vesicular stomatitis virus (VSV) and rabies virus (RV P). We devised a new procedure for combining the results from multiple prediction methods and locating the boundaries between disordered regions and structured domains. To validate the proposed modular organization predicted for RV P and to confirm that the putative structured domains correspond to autonomous folding units, we used two-hybrid and biochemical approaches to characterize the properties of several fragments of RV P. We found that both central and C-terminal domains can fold in isolation, that the central domain is the oligomerization domain, and that the C-terminal domain binds to nucleocapsids. Our results suggest a conserved organization of P proteins in the *Rhabdoviridae* family in concatenated functional domains resembling that of the P proteins in the *Paramyxoviridae* family.

Keywords: phosphoprotein; rabies virus; Sendai virus; intrinsically disordered proteins; protein domains

Introduction

The phosphoprotein is an essential component of the replication machinery of viruses from the *Rhabdoviridae* family, which includes rabies virus (RV) and vesicular stomatitis virus (VSV). P has similar functions in other non-segmented negative strand RNA viruses belonging to the *Paramyxoviridae* (Sendai virus (SV), measles virus, respiratory syncytial virus, and Nipah virus), the *Filoviridae* (Ebola virus and Marburg virus), and the *Bornaviridae* (Borna disease virus), which are grouped together with the *Rhabdoviridae* in the order of the *Mononegavirales* (*MNV*). In all *MNV* viruses, P is part of a set of five common proteins encoded in the same order along the negative strand genome from the 3′ extremity to the 5′ extremity: the nucleoprotein (N), the phosphoprotein (P), the matrix protein (M), the glycoprotein (G) and the RNA-dependent RNA polymerase (L). The nucleoprotein binds to the viral RNA, forming long helical ribonucleoprotein complexes that serve as templates for RNA transcription and replication. Protein P has multiple roles at different stages of the viral cycle, forming a complex with the nascent nucleoprotein that prevents the

*Corresponding author. E-mail address: jamin@embl.fr.

† F.C.A.G. and E.dA.R. contributed equally to this work.

Abbreviations used: RV, rabies virus; VSV, vesicular stomatitis virus; SV, Sendai virus; *MNV*, *Mononegavirales*; N, nucleoprotein; P, phosphoprotein; M, matrix protein; G, glycoprotein; L, RNA-dependent RNA polymerase; aa, amino acid(s); CTD, C-terminal domain; SEC, size-exclusion chromatography; MALLS, multi-angle laser light-scattering; RI, refractometry.

association of the latter with host-cell RNA and forming a two-subunit viral RNA polymerase complex with protein L.

Recent studies have shown that P proteins from viruses of the *Paramyxoviridae* contain long disordered regions that alternate with structured domains,[1–5] and various structural and functional data are available for either VSV P or RV P that suggest the existence of common functional regions. Also, a previous analysis of the amino acid sequence suggested that *Rhabdoviridae* P proteins contain disordered or flexible sequences.[1] The P proteins from both viruses are highly sensitive to proteolysis; domains of RV and VSV P for which the structure were solved by X-ray crystallography were identified by limited proteolysis.[6–8] The N-terminal fragments of both VSV P and RV P interact with N in the N^0-P complex, which is sufficient for maintaining N (referred to as N^0) in a soluble, RNA-free form.[9,10] A structured domain in the central part of VSV P is responsible for the dimerization of the protein,[7] in agreement with our recent demonstration that recombinant, full-length P proteins from both VSV and RV form dimers in solution.[11] The oligomerization domain of RV P was shown to be located in the central region (amino acids (aa) 52–189),[12] although its precise location and boundaries were not clearly determined.[13] The C-terminal domains of RV and VSV P bind to N–RNA complexes, and their structures were solved by X-ray crystallography[6] and NMR,[14] respectively.

In infected cells, apart from the full-length protein, four N-terminally truncated forms of RV P (P2, aa 20–297; P3, aa 53–297; P4, aa 69–297; and P5, aa 83–297) are produced by a ribosomal scanning mechanism.[15] Full-length P and P2 are located in the cytoplasm, whereas P3, P4, and P5 are located in the nucleus. The subcellular localization of the different P forms is regulated by a CRM1-dependent nuclear export signal (NES-1) located between residues 49 and 58 and NES-2 in close association with the nuclear localization signal (NLS) that are both located within the globular C-terminal domain (CTD) (aa 211–214 and 260).[16,17] The C-terminal region (aa 172–297) also contains binding sites for STAT1 and the promyelocytic leukaemia protein (PML), two proteins involved in the interferon-induced antiviral response.[18–20] Finally, the region of aa 139–172 contains a binding site for the cytoplasmic dynein light chain (LC8) that could explain how RV nucleocapsids are transported in neuronal axons towards the cell body.[21,22]

For many years after the first protein structures were determined,[23–25] biological activity of proteins has been associated with folding of the polypeptide chain into a well-defined 3D structure. This idea was supported by the discovery that the biological activity disappears when proteins unfold.[26,27] However, this structure/function paradigm was recently challenged with the discovery that many proteins are fully disordered or contain disordered regions under physiological conditions.[28–31] Also, several proteins were found to exist in a disordered form in isolation but to fold into a well-defined structure upon binding to a partner.[32–36] More surprising was the realization that disorder was required for, or participated in, the biological function of some proteins.[28,31,33,37–41] Several approaches were developed for sorting out disordered proteins from structured proteins on the basis of their amino acid sequence.[29,38,42] Application of such algorithms to whole genomes suggested that as much as one-third of the proteins from eukaryotic organisms contain long disordered regions.[38,42] This raised a new challenge in structural biology for locating disordered regions and folded domains within the protein. Various algorithms were developed for predicting the location of disordered regions from the amino acid sequence (for recent reviews, see Refs. 5 and 43). Some of these algorithms rely on the relationship between the amino acid sequence and the disordered state of the polypeptide chain (SEG, FoldIndex, Globplot, HCA plot, and NORSp).[44,45] For example, it was recognized early that a protein region rich in charged residues and poor in hydrophobic residues is likely to be disordered.[29,44] Other programs rely on trained neural networks anduse sets of known disordered proteins (RONN, DisEMBL, and Disopred2), or combine both approaches (PONDR). However, different methods may yield different predictions and it is difficult to choose the method appropriate to a given protein.

Here, we analyzed the amino acid sequence of the phosphoproteins from SV, VSV, and RV using algorithms available through WEB servers for predicting the location of disordered regions and of secondary structure elements, and we analyzed the amino acid sequence conservation. To avoid having to choose among the prediction methods, we devised simple scoring procedures for integrating the predictions from the different algorithms, and for yielding consensus locations of boundaries between ordered and disordered regions. As test cases, we applied this approach to SV and VSV P for which the location of several structured domains and disordered regions are known from previous structural studies.[2,3,7,14,46] Then, we obtained similar predictions from analysis of the amino acid sequence of RV P. The properties of the predicted structured domains have been tested in a cellular context using a yeast two-hybrid approach, and *in vitro* by producing and characterizing recombinant protein fragments. Circular dichroism was used to determine the secondary structure content and size-exclusion chromatography (SEC) combined with detection by multi-angle laser light-scattering was used to determine the molecular size, molecular mass and the stoichiometry of the different protein fragments. Native gel mobility-shift experiments were used to test the interaction of the various domains with the nucleoprotein–RNA complex, which represents the physiological partner. Altogether, these results provided by the *in vivo* and *in vitro* characterizations of the various P fragments confirmed the *in silico* analysis and allowed us to propose a model for the modular organization of RV P.

Results

Predictions from the amino acid sequence of SV, VSV and RV P

Strategy for predicting the location of disordered and structured regions

We attempted to predict the boundaries between structured and unstructured regions in the SV, VSV and RV P proteins. The SV P was used as a test case, since knowledge about the positions of some of its structured and unstructured domains is available.[2,3,7,8,14,47,48] The location of disordered regions in the sequence of the P proteins was predicted from the amino acid sequence by using 16 different algorithms available through WEB servers (Supplementary Data Figure Fig. S1). MeDor, a recently developed meta-server for disorder prediction, yielded similar pictures (Supplementary Data Fig. S2).[49] The predictions showed significant differences in the number, position and length of the predicted disordered regions. The discrepancies between the results from different algorithms argue for the existence of different types ("flavors") of disorder,[50,51] but raise the problem of choosing the best algorithm and/or strategy for locating the boundaries of disordered regions in a given protein.[43,51]

Recently, we devised a simple voting procedure that combines multiple predictions of disordered regions and predicts the boundaries between structured and unstructured regions. We used it to delimit the nucleoprotein–RNA (N–RNA) binding C-terminal domain of VSV P protein.[14] Here, we improve this approach by including secondary structure predictions and amino acid sequence conservation, and use it to predict the modular organization of SV, VSV and RV P. Our method requires no advised input from the experimentalist. It yields a consensus prediction that integrates the results from multiple disorder predictions, from multiple secondary structure predictions, and from amino acid sequence conservation among homologous proteins. For disorder predictions and secondary structure predictions, we devised scoring methods based on simple binary voting procedures. For each disorder prediction, residues were simply ranked in a binary manner, ordered or disordered, using the default threshold set for each algorithm in the WEB servers. The statistical weights yielded by some of these algorithms were not accounted for, since they were not available for each predicting method and were difficult to compare. For each predicting algorithm, a score of 0 was attributed to each residue predicted to be disordered, and a score of 1 was attributed to each one predicted to be in a structured region. A D-score was calculated by adding the values for each residue and dividing by the number of used algorithms. We arbitrarily defined two threshold levels at 0.75 and 0.50; residues with a D-score <0.50 were assigned as disordered, while residues with a D-score >0.75 were assigned as structured. The structural state of the intermediate zone, with a D-score between 0.50 and 0.75, was considered as ambiguous. Residues with intermediate D-scores were found to occur at the boundaries between ordered and disordered regions, but they could also indicate protein regions that are stabilized mainly by tertiary or intermolecular interactions.

A complementary approach for deciphering the modular organization of a protein consists in identifying structured domains by predicting the location of secondary structures.[43,51] Numerous algorithms, available through WEB servers, predict the location of secondary structure elements on the basis of the amino acid sequence, and some algorithms combine multiple methods.[52,53] Here again, we used a simple voting method for combining predictions from different algorithms and for defining consensus locations of predicted secondary structures (SS-score).

Finally, as mutation rates tend to be higher in disordered regions than in structured domains,[54] structured domains are expected to appear as conserved regions in multiple sequence alignments of homologous proteins. Exceptions could be found in disordered regions interacting with partners and hence presenting some sequence conservation. To score the conservation rate in multiple sequence alignments we used the score provided by the AL2CO software,[55] and we considered regions with a normalized score (C-score) >0.5 as conserved regions. Finally, we defined a global score (G-score), as a non-weighted, linear combination of the scores for disorder prediction (D-score), secondary structure prediction (SS-score) and sequence conservation (C-score). Again, we arbitrarily defined a threshold of 0.5 in order to discriminate between disordered (G-score <0.5) and structured regions (G-score >0.5).

Amino acid sequence analysis of SV P

To assess the reliability of the above-described scoring methods, we analyzed the amino acid sequence of the SV P protein. Figure 1a shows the location of functional regions and Fig. 1f shows the location of secondary structure elements found in high-resolution structures of these two domains.[2,3,46] The structure of the central oligomerization domain (aa 320–429) was solved by X-ray crystallography. It is a tetramer in which each monomer is composed of a long C-terminal α-helix (aa 364–429) that forms a tetrameric coiled coil plus an N-terminal helical bundle of three short α-helices (Fig. 1f).[3] The structure of the C-terminal N–RNA interacting domain (aa 516–568), solved by NMR spectroscopy, consists of a three-helix bundle.[2] NMR also showed that the structured C-terminal domain (aa 516–568) is linked to a flexible domain (aa 474–504) by an even more flexible linker (aa 505–515).[2,46] There is no information on the structure of residues 1–319 or 430–473.

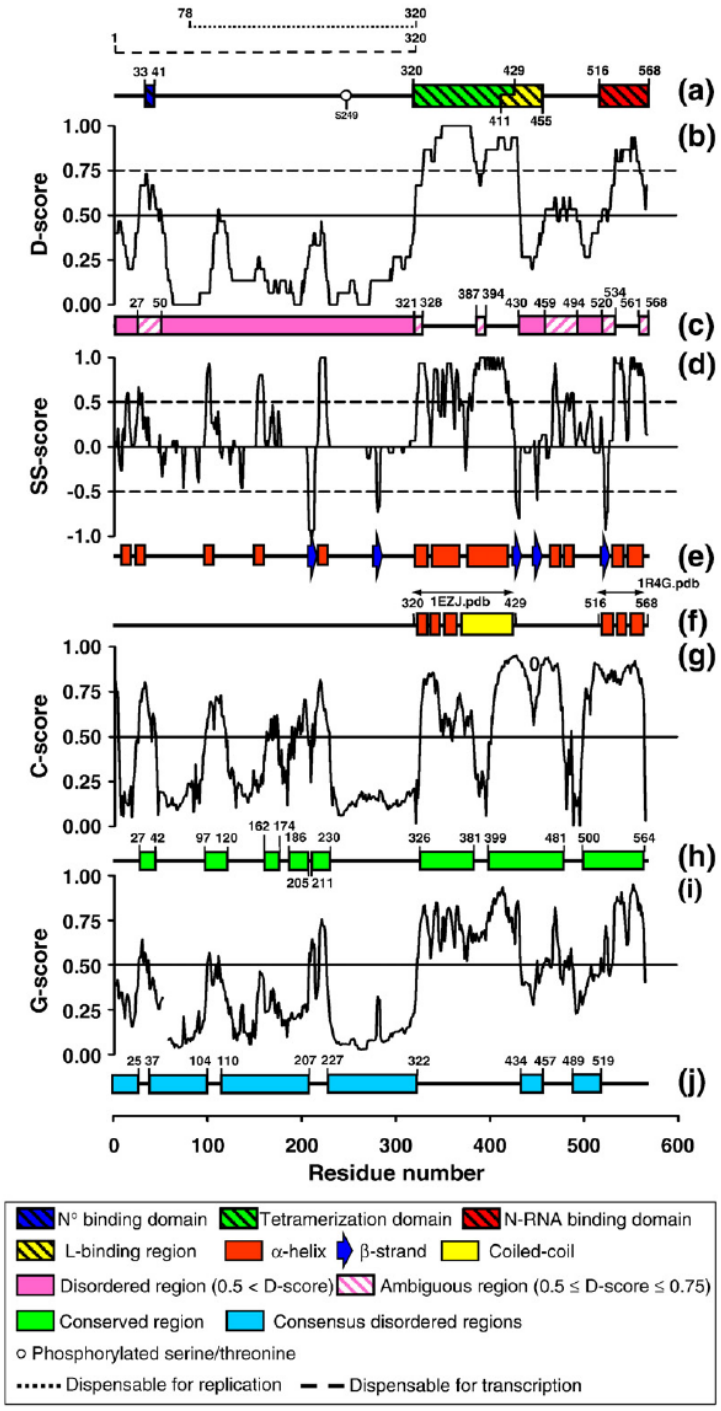


Fig. 1. Predicting boundaries between structured and disordered regions in Sendai virus (SV) P. (a) Location of regions of known function.[47,74,75] The conventions used in this figure are shown in the inset below. (b) Score for disorder (D-score) as a function of residue number calculated as described in the text. (c) Location of consensus disordered regions with a D-score <0.50. Regions with a score in the range 0.50–0.75 are shown as striped boxes. Numbers above the boxes indicate the boundaries of the consensus disordered regions. (d) Score for secondary structure prediction calculated from 15 different predictions (see the text). (e) Consensus location of predicted secondary structure as obtained from the scoring procedure described in the text (SS-score >0.5 or <−0.5, respectively). (f) Location of secondary structure elements in the known 3D structures of two domains,[2,3] with the PDB filenames above the drawing. (g) Score for sequence conservation calculated with the AL2CO software from a multiple sequence alignment obtained with T-Coffee. Conserved residues are defined as regions with a C-score >0.5. (h) Consensus location of conserved regions as defined by the C-score. (i) Global score calculated as a linear combination of the D-score (b) the absolute value of the SS-score (d) and the C-score for sequence conservation (g). Structured regions have a score >0.5 and disordered regions have a score <0.5. (j) Consensus location of predicted disordered and structured domains as defined by the G-score.

Although disorder predictions for SV P displayed significant discrepancies among the various algorithms (Supplementary Data Figs. S1A and S2A), the normalized D-score shown in Fig. 1b revealed the existence of three main disordered regions (score ≤0.5) ranging from residue 1 to residue 321 (with the exception of the 28–49 region), from residue 430 to residue 459, and from residue 494 to residue 520, which are thus considered as the consensus disordered regions (Fig. 1c). The D-score clearly highlighted the location of the two structured domains, namely a central region spanning residues 322–429, which closely corresponds to the tetramerization domain (aa 320–429),[3] and a C-terminal region encompassing residues 521–568, which corresponds to the structured C-terminal domain (aa 516–568).[2]

Predictions of secondary structures (α-helices and β-strands) supported these conclusions. Figure 1d shows the normalized SS-score for α-helix/β-strand

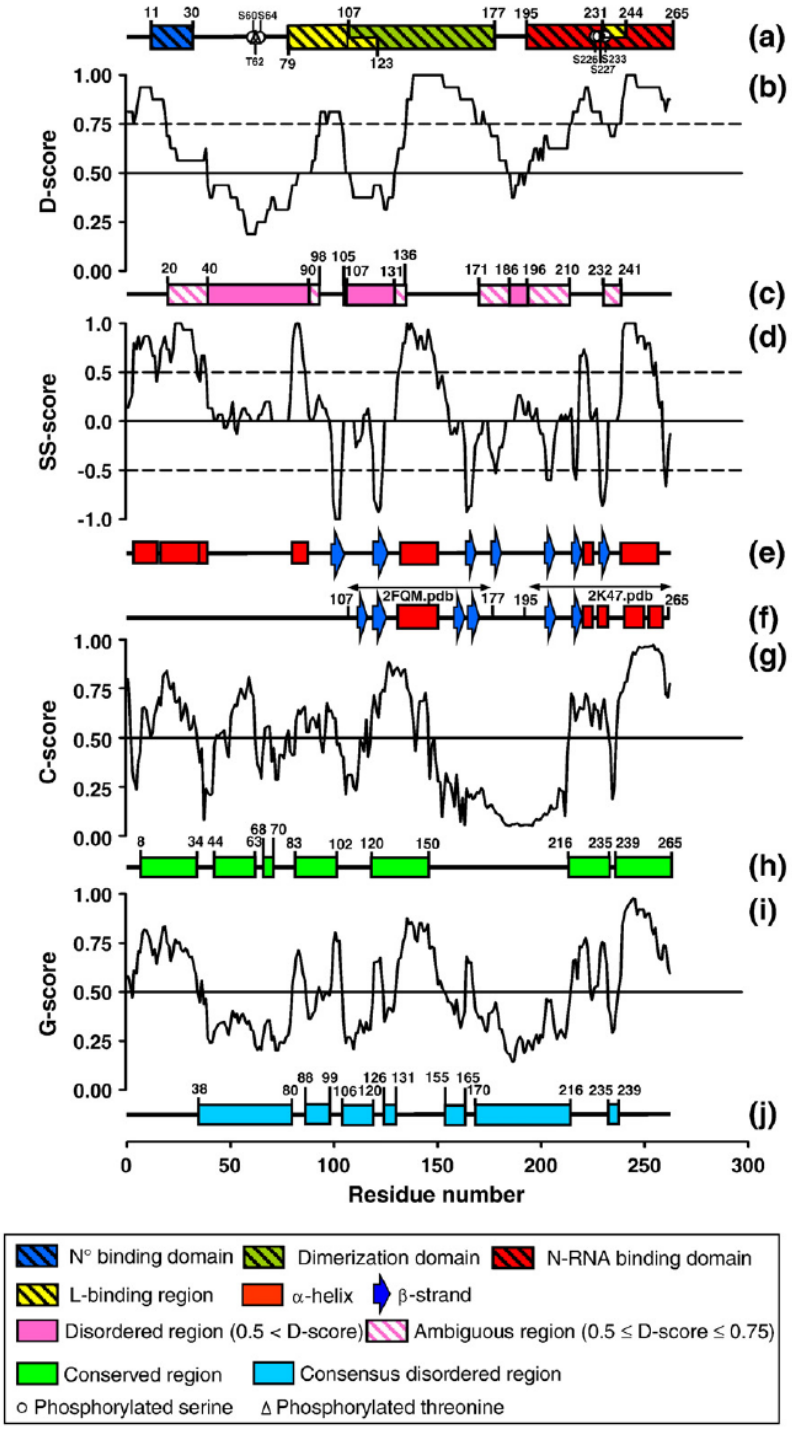


Fig. 2. Predicting boundaries between structured and disordered regions in vesicular stomatitis virus (VSV) P. (a) Location of regions of known function.[7,8,10,58–66] The conventions used in this figure are shown in the inset. (b–j) see the legend to Fig. 1.

prediction and Fig. 1e shows the consensus locations of these secondary structure elements. Within the known structured domains (Fig. 1f),[2,3] secondary structures were predicted correctly, with the exception of the first α-helix in the C-terminal domain that was predicted as a β-strand. This C-terminal domain is rather flexible in isolation and is stabilized upon binding to N–RNA complexes.[56] In particular, residues in the first helix undergo proton exchange, and the helix is in conformational exchange with the unfolded form.[48] The long α-helix predicted in the central domain was also predicted to form a coiled coil (aa 375–420) by the analysis of the HCA plot (see Fig. 2 in Ref. 57), as well as by various coiled coil predictors (see Materials and Methods), although with a level of confidence generally lower than that set by default in the WEB servers (data not shown). The coiled coil of SV P exhibits a different packing

pattern and is more hydrophilic than typical coiled coils,[3] which may explain the failure of these algorithms to clearly predict this structure from the sequence.

A multiple sequence alignment was obtained for P proteins from four members of the *Respirovirus* genus in the *Paramyxoviridae* family. The normalized C-score shown in Fig. 1g indicated five short conserved regions in the N-terminal half of the protein (aa 27–42, 97–120, 162–174, 186–205, 211–230), and three long and more conserved regions in the C-terminal half (aa 326–381, 399–481, 500–564) (Fig. 1h). The three conserved regions in the C-terminal half of P cover most of the tetramerization domain and the entire C-terminal domain involved in binding to N–RNA complexes. However, part of the tetramerization domain is not conserved (aa 382–398) and two regions (aa 430–459 and 500–520) predicted to be disordered were found to be conserved.

Finally, the global score (G-score) shows that the C-terminal half of the protein is predicted to be more structured than the N-terminal half (Fig. 1i and j). The two regions with the highest G-score in the C-terminal half (aa 323–433, 520–566) match the domains of known 3D structure. In addition, our analyses revealed the presence of three regions in the N-terminal half (aa 26–36, 105–109 and 208–226) and one region in the C-terminal half of SV P (aa 458–488) with a G-score >0.5, suggesting that they are structured. Two of these regions (aa 26–36 and 458–488) with intermediate D-score are predicted to contain secondary structure elements (|SS-score| >0.5) and are conserved among homologous proteins (C-score >0.5). One (aa 26–36) corresponds to the binding site of N in the N^0–P complex, while the other (aa 458–488) partially overlaps the flexible region (aa 474–504) identified in the NMR experiments.[2] The two other N-terminal regions (aa 105–109 and 206–226) that have D-scores <0.5 but that clearly stand out from the background (Fig. 1b), are predicted to form secondary structure elements (|SS-score| >0.5) and are conserved among homologous proteins (C-score >0.5). Until now, there is no experimental evidence for the presence of structured domains in these regions. The sequences are short and could correspond to flexible regions in isolation, which may fold upon interaction with viral or cellular partners.

In conclusion, for SV P, our scoring method predicts the location of known boundaries between disordered and structured regions with an accuracy of a few residues and indicates the possible existence of other structured regions.

Amino acid sequence analysis of VSV P

A similar analysis was carried out with the amino acid sequence of VSV P (Indiana serotype) because numerous experimental data about the structure and function of different regions of this protein are available. Figure 2a shows the location of functional regions along the sequence of P,[7,8,10,58–66] and Fig. 2f shows the location of secondary structure elements in the high-resolution structures of the central dimeric domain,[7] and in the C-terminal N–RNA binding domain[14] determined by X-ray crystallography and NMR spectroscopy, respectively. The D-score (Fig. 2b) revealed the presence of three disordered regions (aa 40–90, 107–131 and 186–196) and four structured regions (aa 1–19, 99–104, 137–170 and 211–265 with the exception of 232–241). Figure 2c shows the location of consensus disordered regions. Figure 2d shows the SS-score for α-helix/β-strand prediction, and Fig. 2e shows the consensus location of these secondary structure elements. In the central domain, the α-helix and two β-strands found experimentally in the 3D structure (Fig. 2f)[6] were predicted correctly, but two other β-strands, one on each side of the α-helix were not predicted. In the C-terminal domain, all secondary structure elements were also predicted correctly apart from a short α-helix (aa 229–234), which was predicted as a β-strand. A coiled coil was predicted by four algorithms as well as by analysis of the HCA plot with a consensus location between residues 8 and 36, but these predictions had a level of confidence below the threshold value used in the various WEB servers and are, therefore, not indicated in Fig. 2. Secondary structures were predicted in two regions of the N-terminal part of VSV P. Firstly, α-helices (aa 4–34) are predicted in the region that binds to the N in the N^0–P complex.[10] Secondly, one short α-helix (aa 81–87) and one short β-strand (aa 101–104) are predicted in the region that binds to L.[58]

A multiple alignment of the amino acid sequences of P proteins from 12 members of the *Vesiculovirus* genus revealed an overall 5% identity and 20% similarity. The normalized C-score for sequence conservation calculated from the multiple sequence alignment is shown in Fig. 2g, and boundaries for consensus conserved regions are shown in Fig. 2h. Two clusters of conserved regions (C-score >0.5) are found in the sequence of VSV P. A first cluster (aa 8–34, 44–63, 68–70, 83–102 and 120–150) encompassing the N-terminal and central regions is separated from a second cluster corresponding to the C-terminal N–RNA binding domain (aa 216–235 and 239–265) by a long non-conserved region (aa 151–215). In the first cluster, one conserved region (aa 8–34) matched with the N-terminal N^0 binding region (aa 11–30) and another one (aa 120–150) with the main part of the central dimerization domain (aa 107–177). A third region (aa 44–63) contains the phosphorylation sites (Ser60, Thr62, and Ser64) that control the transcriptional activation,[61,62] and a fourth region (aa 83–102) contains the α-helix and β-strand predicted in the L-binding region.[58]

Finally, the G-score calculated for VSV P suggests the presence of four main structured regions (aa 1–37, 81–105, 121–169 and 217–265) separated by three main disordered regions (aa 38–80, 106–120 and 170–216) (Fig. 2i and j). The predicted N-terminal domain (aa 1–37) encompasses the binding site of N in the N^0–P complex (aa 11–30).[10] The second predicted structured regions (aa 81–105)

contains one of the L-binding sites. The α-helix predicted in this region (Fig. 2d and e) is located in a conserved region (Fig. 2g and h) that is otherwise predicted to be disordered (Fig. 2b and c), suggesting that these secondary structure elements could fold only upon binding to the L polymerase. In the central dimerization domain (aa 121–169), the predictions indicate the presence of three structured regions (aa 121–126, 132–154, and 166–169) that corresponds closely to a large part of the second β-strand (aa 119–126), the middle kinked α-helix (aa 131–157) and the main part of fourth β-strand (aa 165–170) found in the crystal structure.[7] The D-score predicts that the two N-terminal β-strands are disordered (aa 107–131) and our global prediction score fails to identify the first and third β-strands. The last predicted regions (aa 217–265) corresponds to most of the C-terminal N–RNA binding domain (aa 195–265), with the exception of the initial β-strand and the short aa 235–239 region that are predicted to be disordered. The aa 235–239 region corresponds precisely to the place where an addi-

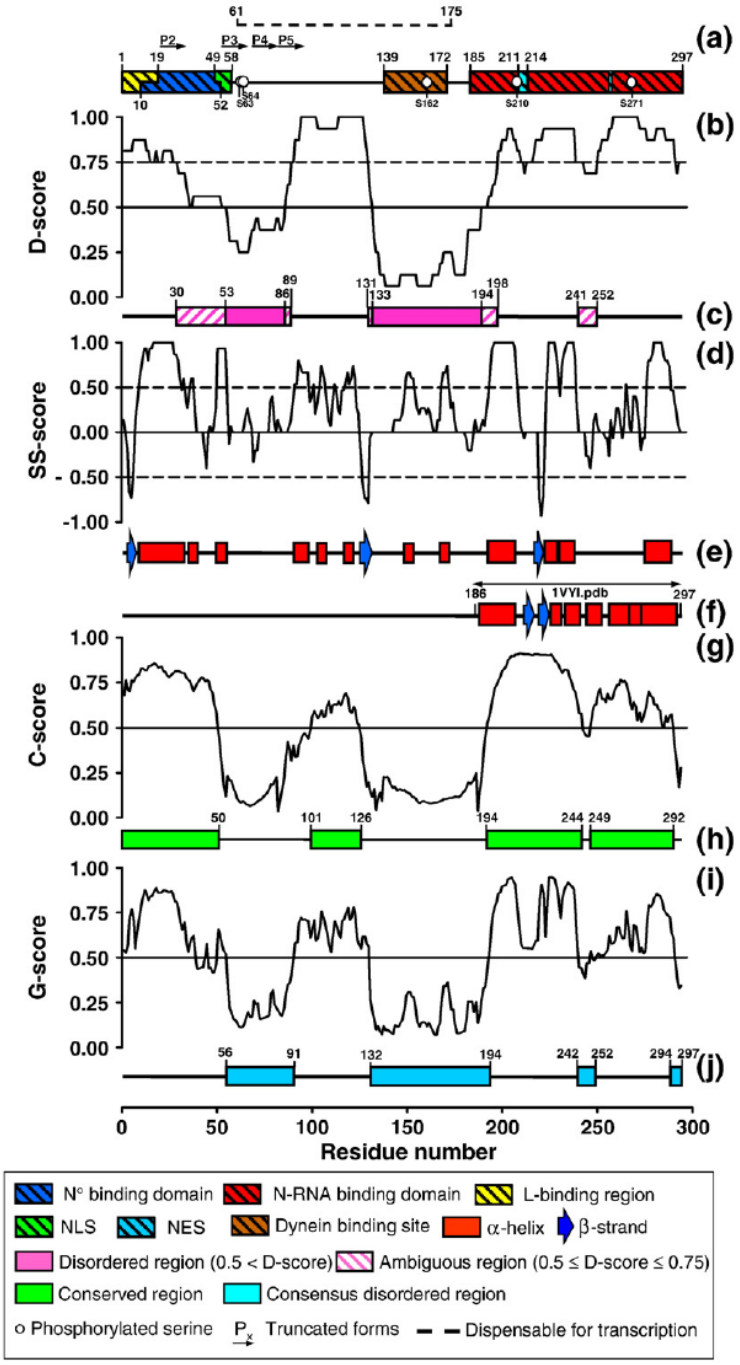


Fig. 3. Predicting boundaries between structured and disordered regions in rabies virus (RV) P. (a) Location of regions of known function,[6,9,12,16,18–22,68] The conventions used in this figure are shown in the inset. (b–j) see the legend to Fig. 1.

tional α-helix is inserted in the C-terminal N–RNA binding domain of RV P.[14]

A similar pattern of disordered and structured regions was found with the amino acid sequence of VSV New Jersey (NJ) serotype (data not shown). P protein from NJ serotypes are longer (274 aa) than those of Indiana (IND) serotypes (265 aa), containing two small gaps at positions 45 and 71–73, and a large insertion of 14 residues between residues 195 and 196 (numbering according to the IND serotype). This latter insertion is predicted to be disordered, resulting in a longer C-terminal disordered region than in the IND serotype.

Amino acid sequence analysis of RV P

Figure 3 shows the results of the same analyses carried out with the sequence of RV P. Figure 3a shows the location of functional regions along the sequence of RV P,[12,13,21,22,67,68] and Fig. 3f shows the location of secondary structure elements in the high-resolution structure of the C-terminal N–RNA binding domain determined by X-ray crystallography.[6] The consensus disordered regions (Fig. 3c) determined from the D-score (Fig. 3b) revealed the presence in RV P of two disordered regions (aa 53–86 and 133–194) and three structured regions (aa 1–29, 90–130 and 200–297 with the possible exception of 241–252). The consensus locations of several of the secondary structures (Fig. 3e) derived from the SS-score (Fig. 3d) indicate the presence of α-helices and β-strands in the three predicted structured regions as well as of two α-helices in the second disordered regions. In the N-terminal region, which binds to N in the N^0–P complex and to L,[9,10,68] the analysis predicted the presence of one β-strand and three α-helices. Within this region, a coiled coil was also predicted by four algorithms and by analysis of the HCA plot with a consensus location between residues 9 and 42, but these predictions had a level of confidence below the threshold value used in the various WEB servers and are, therefore, not indicated in Fig. 3. In the C-terminal region, the analysis predicted most of the secondary structures experimentally found in the three-dimensional structure of the C-terminal domain of RV P (Fig. 3f),[6] with the exception of one β-strand and three α-helices.

A multiple alignment of the amino acid sequences of P proteins from 13 members of the *Lyssavirus* genus revealed an overall 25% identity and 51% similarity. The consensus conserved regions (Fig. 3h) derived from the normalized C-score (Fig. 3g) revealed the presence in the sequence of RV P of four conserved regions (C-score >0.5: aa 1–50, 101–126, 194–244 and 249–292) that are located within or correspond to the otherwise predicted structured regions (D-score >0.5, Fig. 3a).

Finally, the G-score calculated for RV P points out the presence of three structured domains (aa 1–55, 92–131 and 195–293) separated by two disordered regions (aa 56–91 and 132–194) (Fig. 3i and j). The predicted N-terminal domain (aa 1–55) encompasses the binding site of N in the N^0–P complex

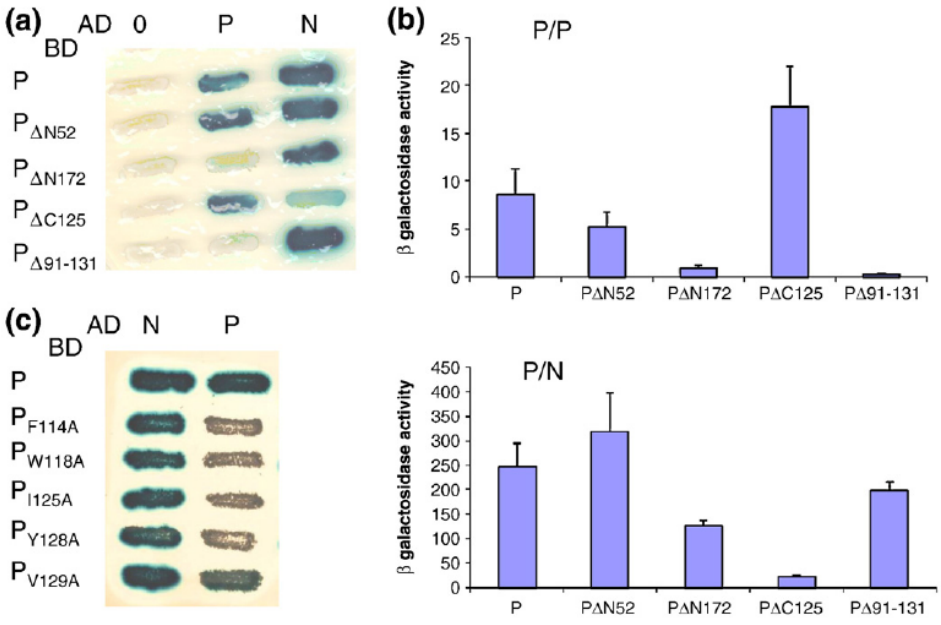


Fig. 4. Binding of RV P to P or N as assessed by a yeast two-hybrid approach. (a and b) Deletion mutants. (c) Point mutations in the central domain. L40 yeast cells were co-transformed with the indicated combinations of plasmids. The interaction between mutants of P fused to the DNA-binding domain (BD) of Lex A (rows in the figure) and P or N genes fused to GAL4 AD (columns in the figure) was assessed by the appearance of blue colonies in the presence of X-Gal (a and c) and by assaying the β-galactosidase activity of yeast grown in liquid medium (b). Quantitative results were obtained from three independent yeast co-transformants assayed with ONPG as substrate. Error bars indicate the standard deviation.

(aa 4–40),[9] the predicted central domain (aa 92–131) could correspond to the dimerization domain, since it is located within the region necessary for oligomerization that was previously identified by deletion (aa 52–189),[12] whereas the predicted C-terminal structured domain (aa 195–293) corresponds closely to the known N–RNA binding domain (aa 186–296). A small region (aa 242–252) in the middle of the C-terminal domain is predicted disordered by the different scores. It corresponds to the end of helix α2, the entire helix α3, and the loop connecting the two helices, a part of the C-terminal domain of RV P that is absent from the homologous C-terminal domain of VSV P.

Experimental analysis of domain structure in RV P

Yeast two-hybrid experiments

Yeast two-hybrid experiments were performed on agar plates and in liquid cultures for quantitative analysis of the interactions (Fig. 4). RV P protein associated with itself and with N (Fig. 4a and b). Deletion mutants were used to determine which region of RV P is responsible for the dimerization of P,[11] and for the interaction with N (Fig. 4a and b). Truncation of the first 172 N-terminal residues ($P_{\Delta N172}$) or of the central region ranging from residue 91 to residue 131 ($P_{\Delta 91–131}$) reduced the association of P with itself considerably, whereas truncation of the 52 N-terminal residues (aa 1–52, $P_{\Delta N52}$) slightly reduced β-galactosidase activity and that truncation of the 125 C-terminal residues (aa 172–297, $P_{\Delta C125}$) led to a twofold increase of β-galactosidase activity (Fig. 4a). These results clearly show that the predicted central domain (aa 92–131 according to the G-score) is critical for the dimerization of RV P as in VSV P.[7] To characterize the dimerization interface, the central region was further dissected. The deletion mutant $P_{\Delta 91–105}$ still associated with P, whereas the mutants $P_{\Delta 106–113}$, $P_{\Delta 114–121}$, and $P_{\Delta 122–131}$ showed a significant reduction of the association with P, indicating that residues 106–131 are required for dimerization (data not shown). In addition, single amino acid mutations were introduced into the central region of P and were tested by using the two-hybrid approach. Five hydrophobic residues were mutated into alanine (F114, W118, I125, Y128, and V129). Although none of these mutations affected the binding to N, they reduced the association with P significantly (Fig. 4c).

Previous data showed that the C-terminal domain of RV P (aa 173–297) binds to N–RNA complexes,[6,59] while the first 40 residues associate with N in the N^0–P complex.[9] In the yeast two-hybrid experiment, truncation of the 125 C-terminal residues of RV P reduced but did not abolish the interactions with N (Fig. 4a and b). The persisting interaction might reflect the interaction of the N-terminal domain of P with N in N^0–P complexes. Truncation of the 172 N-terminal residues or of the central domain (aa 91–131) had little effect on the interactions with N,[67,69] whereas truncation of the 52 N-terminal residues increased β-galactosidase activity slightly (Fig. 4b).

Characterization of recombinant proteins

The full-length P protein and different fragments (P_{NTD}, P_{CED}, P_{CTD}, $P_{\Delta C120}$, $P_{\Delta N52}$, and $P_{\Delta 91–131}$) (Fig. 5) were prepared as recombinant proteins with a C-terminal His_6 tag and purified in a two-step procedure using Ni^{2+} affinity and SEC. $P_{\Delta N52}$ and $P_{\Delta C120}$ were poorly expressed and could be obtained in only small amounts.

The far-UV CD spectrum of full-length P, P_{CED}, P_{CTD}, and $P_{\Delta 91–131}$ exhibited two dichroic bands near 208 nm and 222 nm, indicating the presence of α-helical structure, while the spectrum obtained with P_{NTD} showed one minor dichroic band near 222 nm and one below 200 nm, suggesting a low α-helical content (Fig. 6). The proper folding of the central domain, P_{CED}, was confirmed by 1D NMR spectroscopy (data not shown). Because β-strand and random coil contribute little to the CD signal at

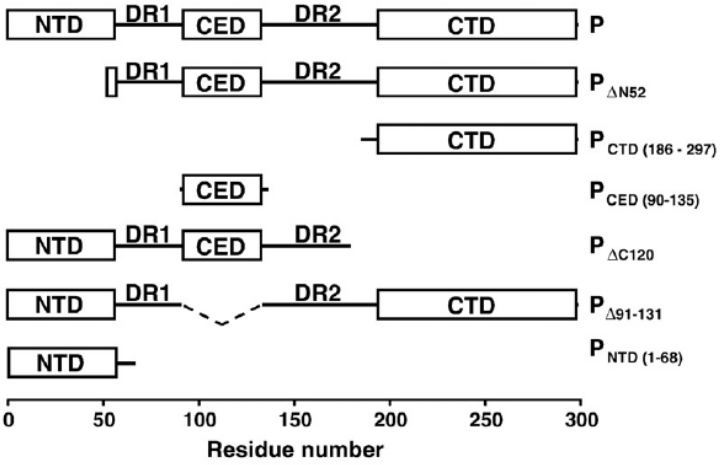


Fig. 5. A schematic representation of the different constructs used for producing recombinant proteins. The predicted disordered regions are labeled DR1 and DR2. NTD, amino-terminal domain; CED, central domain; CTD, carboxy-terminal domain.

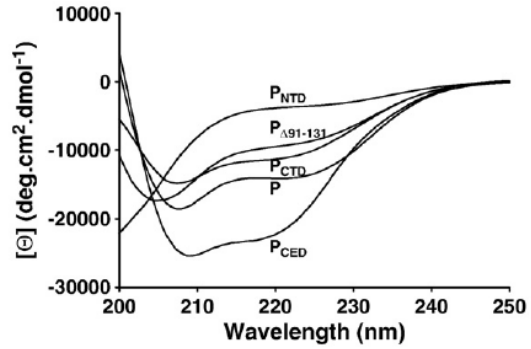


Fig. 6. Circular dichroism spectroscopy of deletion mutants of RV P. CD spectra of different constructs of RV P. Protein spectra were recorded at 20 °C in 20 mM Tri-HCl pH 7.5 and 150 mM NaF.

222 nm, the average amount of α-helix was estimated grossly from the molar ellipticity value at 222 nm.[70] The values for P_{CED} (−21,100 deg cm^2 dmol^{-1}, 28 aa in α-helix), P_{CTD} (−14,000 deg cm^2 dmol^{-1}, 45 aa in α-helix), $P_{\Delta 91-131}$ (−9,200 deg cm^2 dmol^{-1}, 67 aa in α-helix), and full-length P (−11,100 deg cm^2 dmol^{-1}, 94 aa in α-helix) (Fig. 6) corresponded closely to the fraction of α-helical structure predicted from the amino acid sequence of these proteins (P_{CED}, 19 aa; P_{CTD}, 43 aa; $P_{\Delta 91-131}$, 73; and P, 92 aa) (Fig. 3). The value for P_{NTD} (−3600 deg cm^2 dmol^{-1}, 7 aa in α-helix) indicated a lower α-helical content than that suggested by the secondary structure predictions (P_{NTD}, 30 aa) (Fig. 3). The number of residues in α-helical conformation in P_{CED} and in $P_{\Delta 91-131}$ adds up to that calculated for full-length P (28 aa+67 aa=95 aa), supporting the hypothesis that P_{CED} and P_{CTD} are autonomous folding units. However, the number of α-helical residues in P_{CTD} calculated from the CD spectrum (45 aa) was significantly lower than that derived from the crystal structure (67 aa), and summing α-helical residues in P_{NTD}, P_{CED}, and P_{CTD} yields a number of residues (7 aa+28 aa+45 aa=80 aa) significantly lower than that measured for the full-length protein (94 aa). This discrepancy could be accounted for by the presence of additional structure in a region of P other than P_{NTD}, P_{CED} or P_{CTD}. It is possible also that the predicted α-helices in the N-terminal domain fold only in the context of a structure that is larger than the tested fragment of aa 1–68.

Analysis by SEC using on-line detection by multi-angle laser light-scattering (MALLS) and refractometry (RI) indicated that full-length P, P_{CED}, P_{CTD}, and $P_{\Delta 91-131}$ were monodisperse. P_{NTD}, P_{CTD}, and

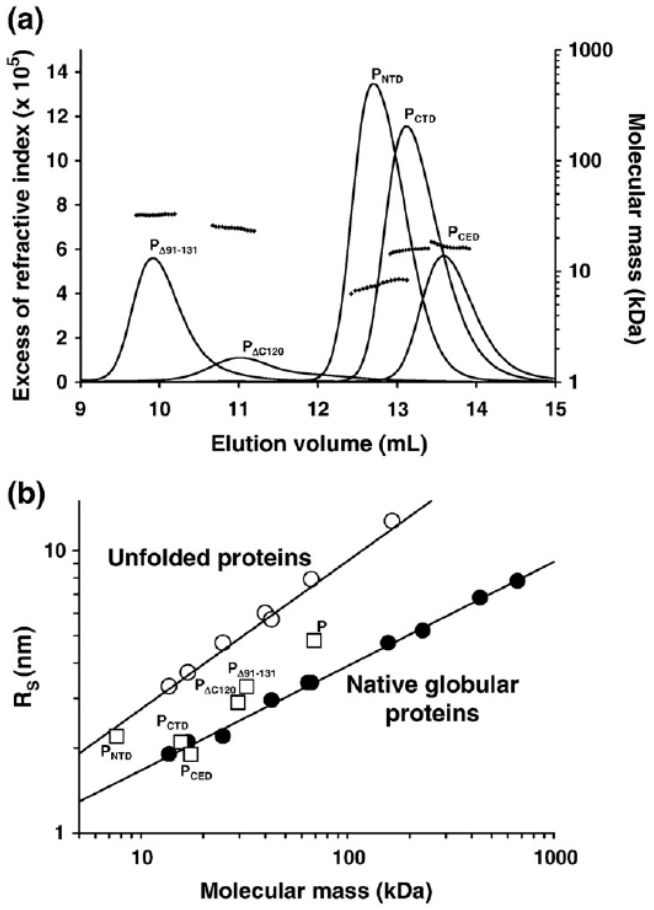


Fig. 7. Molecular mass and size of deletion mutants of RV P. (a) Molecular mass measured by SEC-MALLS-RI. The line shows the SEC elution profile as monitored by refractometry. The crosses show the molecular mass calculated from light-scattering and refractometry data. (b) Plot of Stokes' radius measured by SEC against the molecular mass measured by MALLS-RI. The filled and open circles show data taken from Ref. 71 for standard globular proteins in native or unfolded forms, respectively.

$P_{\Delta 91–131}$ were monomeric with weight-averaged molecular mass of 7.6±0.5 kDa, 15.6±0.2 kDa, and 32.5±0.3 kDa (Fig. 7a), respectively, in agreement with the expected mass calculated from the amino acid sequence (P_{NTD}, 8696 Da; P_{CTD}, 13,989 Da; and $P_{\Delta 91–131}$, 29,306 Da). P_{CED} and P form dimers with an average molecular mass of 14.7±0.2 kDa (calc. 6501 Da) and 69±2 kDa (calc. 33,616 Da), respectively.[11] The hydrodynamic radii (Stokes' radius, R_S) of full-length P, P_{NTD}, P_{CED}, P_{CTD}, $P_{\Delta C120}$, and $P_{\Delta 91–131}$ were determined by calibrating the SEC column with standard proteins.[71] On a plot of the hydrodynamic radius as a function of molecular mass (Fig. 7b), P_{CED} and P_{CTD} appeared as native globular proteins, $P_{\Delta 91–131}$, $P_{\Delta C120}$, and full-length P were intermediate between native and unfolded proteins, indicating that these proteins are non-globular, and P_{NTD} appeared as an unfolded protein. A previous analysis by small-angle neutron scattering revealed that full-length P forms elongated dimeric molecules.[11] These experiments confirmed that the predicted central domain of P constitutes the dimerization domain and, together with the CD experiments, that P_{CED} and P_{CTD} fold as independent domains.

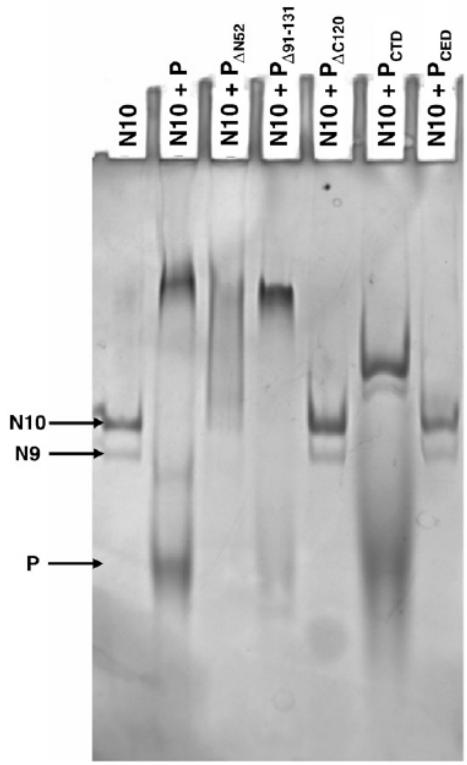


Fig. 8. Binding of deletion mutants of P to nucleoprotein–RNA complexes (N–RNA) using native gel electrophoresis. Recombinant N–RNA rings containing 10 N protomers purified from insect cells were incubated with the different constructs of RV P and loaded onto a native gel. Lane 1, N_{10}-RNA circular complex containing a small amount of contamination with N_9-RNA complex. Lanes 2–7, Interaction of N_{10}-RNA with different deletion mutants of RV P as indicated above the lanes. Proteins were visualized by staining with Coomassie brilliant blue.

Binding to recombinant N–RNA rings

The binding of full-length P and its deletion mutants to a recombinant circular N–RNA complex containing 10 nucleoprotein protomers (N_{10}–RNA)[72] was tested by native gel electrophoresis (Fig. 8). Lane 1 shows the migration of N_{10}–RNA, with a contaminating band of N_9–RNA. In the presence of excess full-length P, a new band appeared corresponding to the complex, while the band corresponding to N_{10}–RNA alone disappeared, and the excess P appeared at the bottom of the gel. When N_{10}–RNA was incubated with $P_{\Delta 91–131}$ or with P_{CTD}, similar shifts were observed for the complex, indicating that, like full-length P, both protein fragments bound to N_{10}–RNA (Fig. 8 lanes 4 and 6). Conversely, when N_{10}–RNA was incubated with $P_{\Delta C120}$ or P_{CED}, the migration of N_{10}–RNA was not affected, indicating no association (Fig. 8 lanes 5 and 7). These results demonstrate that binding of P to N–RNA complexes requires only the presence of the C-terminal domain and is independent of the presence of P_{CED}, and thus, of the dimerization of P.

Discussion

Assessing the scoring procedure

Predicting the location of disordered regions in a protein from its amino acid sequence is an important challenge in structural biology. Many proteins contain disordered regions, which prevent structure determination by X-ray crystallography, although these regions are often found to participate in biological activity. Here, we have devised a simple method for predicting the location of boundaries between structured and disordered regions, which combines different disorder predictions, but which also takes into account secondary structure predictions and sequence conservation. Such a voting procedure integrates information from a large number of algorithms and avoids delicate choices between prediction methods. Clearly, this approach is simplistic and accounts neither for the specificity of the different algorithms nor for the confidence level associated with each amino acid by some algorithms. It may underscore some peculiar types of disorder predicted by some algorithms. Our selection of algorithms is not exhaustive, and the threshold level at 0.5 for defining disorder was chosen arbitrarily. Among the different algorithms used for predicting disordered regions, GlobPlot, PreLink, and the three DisEMBL predictors yielded the most divergent results (Supplementary Data Fig. S1). However, removal of these predictions from the score calculation had little effect on the overall location of disordered regions and modified the location of boundaries by only a few residues (data

not shown), demonstrating the robustness of the method. Therefore, despite possible objections, this procedure has predicted a cutting out of the phosphoproteins from SV, VSV and RV into disordered regions and structured domains that is in agreement with available experimental data, and it has proved successful in predicting the boundaries of folded domains in VSV P[14] and in RV P (this study). Predictions for SV and RV P seemed more clear-cut than those for VSV P, emphasizing that such predictions might be complicated by the existence of regions with poor intrinsic stability that fold upon binding to partners or conserve an enhanced flexibility within a structured domain. Indeed, predictions in the central region of VSV P indicated a complex pattern of short disordered and structured regions, and parts of the domain solved by X-ray crystallography were predicted to be disordered (aa 106–120, 126–131, 155–165).[7] However, a part of this region could fold only upon binding to the L polymerase, and the predicted disordered regions that are folded in the crystal structure exhibit high temperature factor values, which could reflect their dynamical properties.[7] In addition, our analyses have predicted the existence of previously unsuspected structured regions in SV and VSV P, which can now be studied.

A common modular organization within Rhabdoviridae and Paramyxoviridae phosphoproteins

All *MNV* viruses encode a phosphoprotein that has similar roles in the replication cycle, but no similarity is found between P sequences from different *MNV* families or even between P sequences within members of the *Rhabdoviridae*. In the *Paramyxoviridae*, the 3D structure of the C-terminal domain of P is conserved between Sendai and measles viruses,[2,73] and the sequence of the oligomerization domain is well-conserved, suggesting a common homotetrameric coiled coil.[3,74] In the *Rhabdoviridae*, we showed recently that, despite a lack of sequence similarity, the structure of the C-terminal domain is conserved between VSV and RV P.[6,14] However, this structure is different from that of the C-terminal domain of P from Sendai and measles viruses, and the structure of the central dimerization domain of VSV P, an α/β domain,[7] is different from the tetrameric coiled-coil domain found in Sendai virus P.[3]

Despite this lack of similarity, the available structural and functional information combined with a computational analysis of the amino acid sequences of VSV and RV P suggest a modular organization similar to that proposed recently for P proteins from members of the *Paramyxoviridae*, with three structured domains alternating with two disordered regions (Fig. 9).[1]

An N-terminal domain with a potential to form α-helices

The 40 to 50 N-terminal amino acids of VSV and RV P constitute an N-terminal hydrophobic region with α-helical potential, although the far-UV circular dichroism spectrum obtained with the P_{NTD} fragment of RV P (aa 1–68) and the hydrodynamic radius revealed a disordered domain with a low α-helical content in solution. This discrepancy between predictions and experiments might reflect a disordered state in isolation with folding induced upon binding to a physiological partner. In this regard, it is noteworthy that in VSV and RV P, the N-terminal domain contains a binding site for N^0,[9,10] and that in RV P it also contains a binding site for the L protein (aa 1–19).[68] The N-terminal region was previously proposed to constitute the oligomerization domain of RV P, but the first 52 residues were found to be dispensable for dimerization.[13] Here, we confirm that deletion of the first 52 residues of RV P has no effect on the ability of P to self-associate, thus ruling out a role of this region in P dimerization. In P from SV, a small N-terminal domain (aa 1–60), is also predicted to adopt an α-helical conformation, and to be involved in binding to N^0,[1,75] supporting a conserved role for this region of the protein in both *MNV* families.

An N-terminal disordered region

The aa 38–80 region in VSV P (Fig. 2j) and the 56–91 region in RV P (Fig. 3j) are predicted to be disordered. In RV P, a large part of this region (from aa 61) is

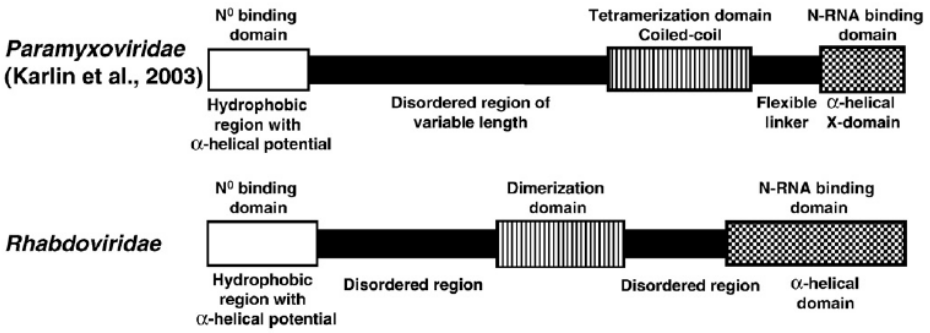


Fig. 9. Modular organization for *Rhabdoviridae* and *Paramyxoviridae* phosphoproteins. Structured regions are represented by wide boxes. The model for *Paramyxoviridae* P protein is reproduced from Ref. 1.

dispensable for transcription,[12] highlighting another similarity with the *Paramyxoviridae*, in which the corresponding disordered region (aa 42–319, Fig. 1a) is of variable length and dispensable for both transcription and replication.[76] In VSV P, this region has an unusually high content of acidic residues (net charge of 15 negative charges) and contains phosphorylation sites (Ser60, Thr62, and Ser64, Indiana serotype) for cellular kinases that control transcription during the viral replication cycle, providing a rational for the conservation of this disordered region among *Vesiculoviruses* (Fig. 2g and h). [61–63,66–78] In SV and RV P, phosphorylation sites for cellular kinases are also found in the corresponding region but, so far, no role in the regulation of viral replication has been found.[13,79–82]

A central oligomerization domain

Recently, we have shown that RV P and VSV P form dimers in solution,[11] and here we demonstrate that the dimerization is mediated by the central domain in both proteins. The dimerization domain of VSV P (aa 107–170) was identified by limited proteolysis and its structure was solved by X-ray crystallography.[7] It is formed by a kinked α-helix flanked on each side by two-stranded β-sheets and assembles into dimers through the packing of the α-helices, thereby forming two domain-swapped β-sheets.[7] In RV P, the predicted dimerization domain appears shorter than that in VSV P. It is expressed in *Escherichia coli* as a soluble, structured protein domain, which is an autonomous folding unit. Secondary structure predictions suggest a different structural organization in RV P with respect to VSV P, with three short α-helices and one β-strand in each monomer (Fig. 2). Experiments with a minireplicon system revealed that the central region of RV P (aa 61–175), and thus the dimerization of P, are dispensable for transcription and replication,[12] precluding a mechanism where P dimers cartwheel along the nucleocapsid.[76] Our predictions reveal the existence of a second structured region (aa 81–87, 100–105) in the center of VSV P, which could be involved in binding to the L polymerase.[58] In *Paramyxoviridae*, the central folded domain forms tetramers through coiled coils,[3] and the tetramerization domain is required for both transcription and replication.[76]

A C-terminal disordered region

A second disordered region is predicted in the C-terminal part of both VSV and RV P. In RV P, this region contains a binding site for the cytoplasmic dynein light chain (LC8).[21,22] This region is considered as a flexible linker in P proteins from the *Paramyxoviridae*.[1,2]

A C-terminal N–RNA binding domain

This domain binds to N–RNA complexes,[6] (and this study) and to proteins from the interferon-induced cellular response, including the STAT1 factor,[18,20] and the promyelocytic leukemia protein[19]. The structure of the C-terminal domain of RV P (aa 186–297), which was solved by X-ray crystallography,[6] is conserved among *Rhabdoviridae*, as shown recently by comparison with the structure of the C-terminal domain of VSV P.[14] A short structured C-terminal domain that bound to the N–RNA complex was also found in Sendai and measles viruses.[56,73]

Disordered regions and functions of P proteins

Long disordered regions are found in many functional proteins,[29,83] and have been associated with specific functions.[31,33] In particular, disordered regions have major roles in exposing sites of post-translational modifications, such as phosphorylation,[31,40] and in promoting multi-molecular assemblies and interactions with different partners.[31,33] The presence of long disordered regions in P proteins of members of the *Rhabdoviridae* family and of other *MNV* families may thus have functional implications. The transcriptional activity of VSV P is regulated by phosphorylation, and residues with a major role in this regulation are located in the long N-terminal disordered region.[61,66,84–86] Phosphorylation sites are found also in the corresponding disordered region of SV and RV P, although no role could be assigned to these modifications in these two viruses.[13,79–82] The phosphoprotein also has a key role in the assembly of the replication/transcription complex by binding to the L-polymerase,[58,68] and to the nucleoprotein in both the N^0–P complex and the N–RNA complex,[9,67,87,88] as well as to cellular partners in the case of RV P.[18–22] Intrinsically disordered proteins can have a role in the assembly of multi-molecular complexes by simply bridging partners together, by regulating the spacing between binding sites, by providing large binding surface or multiple contact points for a large protein, by adopting different conformations in response to different stimuli, or by increasing the speed of interaction through a "fly-casting" mechanism.[31,89] Binding sites for viral (L polymerase for VSV P) or cellular partners (dynein light chain LC8 for RV P) are also found in regions predicted to contain secondary structure elements but otherwise predicted to be disordered, suggesting that these regions could fold upon binding to their partners. Folding upon binding is accompanied by a large decrease in conformational entropy and can provide the high level of specificity associated with a moderate binding strength.[33] The importance of disordered regions in the assembly of the replication complexes of *MNV* viruses was nicely highlighted in the structure of the complex formed between the C-terminal domain of measles virus P protein and a peptide corresponding to a disordered region of N protein.[5,56,90]

In conclusion, in the modular organization proposed for the P protein from *Rhabdoviridae* and *Paramyxoviridae* (Fig. 9),[1] the structured domains share

similar functions, although the structures of some of the domains are different. The presence of disordered regions, however, seems well conserved and could provide the flexibility required for proper function of the proteins; for example, allowing the interaction between different partners within the replication complex. The presence of disordered regions and the function of the structured domains appear to be conserved during evolution.

Materials and Methods

Amino acid sequence analysis

The location of disordered regions within SV, VSV and RV P was predicted by submitting the amino acid sequences to 16 different algorithms accessible through WEB servers. Foldindex was run with a window size of 51 residues.[44] With the different PONDR predictors XL3, VL-XT, XL1-XT, and VSL1, residues with a score >0.5 were considered disordered.[91,92] The NORSp server was used with a window size for disorder set to 25 residues, a structure content cut-off of 12% and with the number of consecutive exposed residues set to 10.[93] The IUPred predictor for long disordered regions was used, and residues with a score >0.5 were considered disordered.[94] The DISOPRED server was run with the rate threshold for predictions of false positive set at 2%.[42] The DisEMBL (Loops/Coils, Hot-loops and Remark465),[95] the RONN,[96] and the PreLink[97] servers were run using the default parameters. Program SEG was used for analyzing sequence complexity.[98] HCA plots were drawn with DRAWHCA,[99] and disordered regions of >20 residues were identified by visual inspection. A simple scoring procedure was used to define consensus disordered regions. For each prediction, residues predicted to be part of a disordered region were assigned a score of 0 while other residues were assigned a score of 1. A D-score for each amino acid was calculated by adding the values from all predictions and normalizing the sum. Consensus disordered regions were defined as regions with a normalized score ≤ 0.50, and structured domains were defined as regions with a normalized score ≥ 0.75.

Similar scoring procedures were used for analyzing predictions of secondary structures. Secondary structure predictions were obtained from 15 different algorithms freely available through WEB servers: DPM,[100] DSC,[101] GOR IV,[102] HHNC,[103] PHD,[104] PREDATOR,[105] SIMPA96,[106] SOPM,[100] SOPMA,[100] NNPREDICT,[107] PORTER,[108] JPRED,[109] PSIPRED,[110] SSPRO,[111] and PROF.[112] All algorithms were run using the default parameters. For each prediction, residues predicted to be in α-helix conformation were assigned a score of 1 while residues predicted to be in extended conformation were assigned a score of −1. An SS-score was calculated for each amino acid by adding values from each prediction and normalizing the sum. α-Helices were defined as regions with an SS-score >0.5 and β-strands as regions with a SS-score <−0.5. Predicted locations of coiled coils were obtained from four different WEB servers: COILS,[113] Paircoil2,[114] Multicoil,[115] and Marcoil.[116]

Multiple sequence alignments were obtained for P sequences from four different *Respiroviruses*: Sendai virus (BAC79134), human parainfluenza viruses 1 (NP_604435) and 3 (BAA00031), bovine parainfluenza virus 1 (NP_037642), from 13 different *Lyssaviruses*: rabies viruses strains CVS-11 (P22363) and Pasteur (P06747), Australian bat lyssavirus (Q9QSP3), Australian bat lyssavirus (Q8JTH2), European bat lyssaviruses 1 (A4UHP9) and 2 (A4UHQ4), Duvenhage virus (O56774), Lagos bat virus (O56773), Mokola virus (P0C569), Aravan virus (Q6X1D7), Irkut virus (Q5VKP5), Khujand virus (Q6X1D3), West Caucasian bat virus (Q5VKP1), or from 12 different *Vesiculoviruses*: VSV strains Indiana Orsay (ACK77581), Indiana San Juan (P03520), Indiana 94GUB Central America (Q8B0H3), Indiana Glasgow (P04879), Indiana Mudd-Summers (P04880), Indiana 98COE North America (Q8B0I3), Indiana 85CLB South America (Q8B0H8), New Jersey Missouri subtype Hazelhurst (P04878), New Jersey Ogden subtype Concan (P04877), Piry virus (Q01769), Chandipura virus (strain I653514) (P16380), and Isfahan virus (Q5K2K6). The multiple alignments were performed with T-Coffee using the default parameters of the server.[117] The C-score characterizing the sequence conservation was obtained for each residue with the AL2CO software,[55] using the Henikoff-Henikoff frequency estimation method, a sum of pairs conservation measure, an averaging window of 20 residues and a gap fraction of 0.5. The C-score corresponding to the standard deviation from the mean was normalized into values ranging from 0 to 1 according to:

$$\text{for } Z-\text{score} \geq 0,\quad 1 - 0.5 \,.\, \exp(-Z\,\text{score}) \tag{1}$$

$$\text{for } Z-\text{score} < 0,\quad 0.5 \,.\, \exp(Z\,\text{score}) \tag{2}$$

A global G-score for locating boundaries between disordered and structured regions was calculated as a linear combination of the D-score for disorder prediction, the absolute value of the SS-score for secondary structure prediction, and the C-score for sequence conservation according to:

$$G-\text{score} = \frac{D-\text{Score} + |SS-\text{Score}| + C-\text{score}}{3} \tag{3}$$

Plasmid construction

The cDNAs encoding full-length P, P deleted of the first N-terminal 52 residues ($P_{\Delta N52}$), P deleted of the first N-terminal 172 residues ($P_{\Delta N172}$), and P deleted of the last C-terminal 125 residues ($P_{\Delta C125}$) were fused to the sequence encoding the DNA binding domain (BD) of Lex A and cloned into pLex (Clontech) as described.[21] The constructs pLex encoding deleted P proteins $P_{\Delta 91-105}$, $P_{\Delta 106-113}$, $P_{\Delta 114-121}$, $P_{\Delta 122-131}$, and $P_{\Delta 91-131}$ were created by deletion in the full-length P gene by using PCR gene fusion. Amino acid substitutions in pLex P (P_{F114A}, P_{W118A}, P_{I125A}, P_{Y128A}, and P_{V129A}) were generated by a two-step, PCR-based, site-directed mutagenesis approach as described.[16] The plasmid pGAD-P was obtained by fusing the full-length P gene with the GAL4 activation domain (AD) of pGAD (Clontech). The plasmid pGAD-N has been described.[21]

Plasmids for expressing full-length RV P (CVS strain) and several fragments (P_{NTD}, aa 1–68; P_{CED}, aa 91–134; P_{CTD}, aa 186–297; $P_{\Delta 91-131}$, $P_{\Delta N52}$, and $P_{\Delta C120}$) in bacteria were constructed by cloning the gene or the corresponding fragment into pET22b(+) using PCR. For each construct, a His_6 tag and a two amino acid linker (Glu-

Leu) were introduced at the C-terminal extremity. All plasmids were verified by standard dideoxy sequencing.

Two-hybrid system

Yeast L40 cells were co-transformed with the plasmids encoding wild-type P or N fused to the AD of GAL4 and the P mutants fused to the DB of Lex A. The interaction was tested by measuring the β-galactosidase activity of histidine-positive clones on plates and in liquid assays. On plates, an X-Gal mixture containing 0.5% (w/v) agar, 0.1% (w/v) SDS, 6% (v/v) dimethylformamide and 0.04% (w/v) X-Gal (5-bromo-4-chloro-3-indolyl-β Δ-galactosidase) was overlaid on freshly transformed cells grown on $Trp^{-}Leu^{-}$ dishes and blue clones were detected after 60 min–18 h at 30 °C. For the liquid assay, cultures were grown overnight and assayed for β-galactosidase activity with *o*-nitrophenylgalactoside (ONPG) as substrate.[118] β-Galactosidase activity was calculated as:

$$\text{Activity} = (1000\ A_{420})/(A_{600}TV) \quad (4)$$

where A_{420} is the absorbance of the reaction mixture at 420 nm, A_{600} is the cell density of the culture (measured as absorbance at 600 nm), T is the reaction time (in min) and V is the volume (in ml) used for the assay.[118]

Production and purification of RV P and its different fragments

The plasmids for expressing full-length P or its fragments were transformed into *E. coli* strain BL21 (DE3) or *E.coli* BL21(DE3)-RIL for expressing P_{CED}. Cells were grown in LB medium containing 100 μg mL^{-1} ampicillin at 37 °C until A_{600} reached 0.6–0.8 a.u. At this point, IPTG was added to a final concentration of 0.5 mM and the temperature was shifted to 24 °C. After an incubation for 5 h, cells were harvested by centrifugation and suspended in buffer A (20 mM Tris–HCl pH 7.5, 150 mM NaCl) supplemented with anti-proteases (Complete™ Protease Inhibitor Cocktail Tablets, Roche), DNAse I (Sigma) and 20 mM $MgSO_4$, and cells were disrupted by sonication. The extract was centrifuged at 20,000 *g* for 1 h at 4 °C. The supernatant was filtered (0.45 μm pore size) and loaded onto a Ni^{2+} resin column, pre-equilibrated with buffer A. The resin was washed with three bed volumes of buffer A, then with three bed volumes of 20 mM Tris–HCl, 1.5 M NaCl, 10 mM imidazole at pH 7.5, and the protein was eluted using buffer A supplemented with 400 mM imidazole. The recombinant protein was loaded onto an SEC column (HiLoad 16/60 Superdex 75 prep grade, GE Healthcare) equilibrated with buffer A. For P_{CED}, the SEC was performed in 20 mM Tris–HCl, 500 mM NaCl at pH 8.5. Separations were performed at a flow rate of 1.0 $mL.min^{-1}$. The purified protein was concentrated (Centricon, Amicon, 10,000 MWCO) and store at 4 °C. The purity of the protein samples were assessed by SDS-PAGE. The identity and integrity of the proteins were confirmed by electrospray mass spectrometry. The experiments were performed on a Quattro II mass spectrometer (Micromass, Altricham, UK) by continuously injecting the sample with a Type 22 pump (Harvard Apparatus) at a flow rate of 5 $\mu L.min^{-1}$. Protein concentrations were measured by spectrophotometry using extinction coefficients calculated from the amino acid sequence.[119]

Circular dichroism spectroscopy

A JASCO J-810 CD spectropolarimeter equipped with a temperature-controller (Peltier system) was used to record the far-UV CD spectra at 20 °C, which were accumulated at least five times. Full-length RV P or fragments thereof were diluted to final concentrations ranging from 10–40 μM in 20 mM Tris–HCl pH 7.5 containing 150 mM NaF using a cuvet with a path-length of 1 mm.

After subtracting the blank signal, the CD signal (in mdeg) was converted to mean molar residue ellipticity (in $deg.cm^2\ dmol^{-1}$) using the following equation:

$$[\theta] = \frac{0.1\,(CD\,signal)}{l\,C\,N_{aa}} \quad (5)$$

where l is the pathlength of the cuvette (in centimeters), C is the protein concentration (in molarity) and N_{aa} is the number of amino acids. The ellipticity at 222 nm was converted to fraction helix values using:

$$f_\alpha = \frac{[\theta]}{[\theta]_{helix}} \quad (6)$$

where $[\theta]_{helix} = -35{,}000$ deg cm^2 $dmol^{-1}$ represents the mean molar residue ellipticity for 100% helix calculated from the following expression derived for peptides:[120]

$$[\theta]_{helix} = (-44000 + 250T)\left(1 - \frac{n}{N_r}\right) \quad (7)$$

where T is the temperature (in °C), n is the number of non-H-bonded CO groups in the peptide and N_r is the chain length in residues. For proteins, we used $n=1$, assuming most CO groups at the C-terminal extremity of α-helices are capped and $N_r=10$ as an average helix length.

SEC-MALLS-RI

SEC was performed with an S75 Superdex column (GE Healthcare) equilibrated with buffer A. Separations were performed at 20 °C with a flow rate of 0.5 mL min^{-1}. Typicallly, 50 μL of a protein solution at a a concentration of 5-10 $mg.mL^{-1}$ was injected. On-line MALLS detection was performed with a DAWN-EOS detector (Wyatt Technology Corp., Santa Barbara, CA) using a laser emitting 690 nm. Data were analyzed and weight-averaged molar masses (*M*w) were calculated using ASTRA software (Wyatt Technology Corp., Santa Barbara, CA) as described.[11]

The excluded (V_0) and total volumes (V_t) were measured with blue dextran and thymidine, respectively. The partition coefficients (K_{av}) were calculated as:

$$K_{av} = \frac{V_e - V_0}{V_t - V_0} \quad (8)$$

where V_e is the elution volume of the protein. The column was calibrated using proteins of known Stokes' radius (R_S) and molecular mass (M_w):[35] bovine serum albumin, R_S 3.4 nm, M_w 67.0 kDa; RNAse A, R_S1.9 nm, M_w=13.7 kDa; ovalbumin, R_S 3.0 nm, M_w 43.5 kDa; β-lactoglobulin, R_S 2,7 nm, M_w 6.8 kDa; and chymotrypsinogen, R_S 2.3 nm, M_w 25 kDa.[35]

Nucleocapsid binding assay

The interaction between circular N–RNA complexes and the various P proteins was investigated by native gel electrophoresis. Circular N–RNA complexes containing 10 nucleoprotein protomers (N_{10}–RNA) were purified as described.[72] The N_{10}–RNA complex was incubated in the presence of full-length RV P, $P_{\Delta 91-131}$, $P_{\Delta N52}$, $P_{\Delta C120}$, P_{CED} or P_{CTD} at 20 °C for 10 min. A native gel of 4% (w/v) polyacrylamide and with a 19:1 (w/w) acrylamide/bisacrylamide ratio was pre-run for 30 min at 200 V in the cold room. Typically, 10 μl of protein mixture at about 1 mg/ml (where P fragments were in molar excess with respect to N–RNA) was loaded, and the gel was run for 3 h at 200 V. Complexes were identified by staining with Coomassie brilliant blue.

Acknowledgements

This work was supported by the interdisciplinary program Maladies Infectieuses Emergentes from the CNRS, and by a grant from the French ANR (ANR-07-001-01 (ANRAGE)) and Lyonbiopôle. We thank Bertrand Lefèbvre, Denis Daveloos and Bernard Brasmes (CRSSA, Grenoble) for mass spectrometry measurements, and Baptiste Fouquet and Hélène Raux for the amino acid substitutions in P and the test in the two-hybrid system. We thank Noël Tordo for extensive discussions and critical reading of the manuscript. E.A.R. was supported by postdoctoral fellowships from the University Joseph Fourier and from the ANR program (ANR-06-JCJC-0126-01). F. C.A.G. and C.L. were supported by MENRT fellowships from the French government.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.03.061

References

1. Karlin, D., Ferron, F., Canard, B. & Longhi, S. (2003). Structural disorder and modular organization in Paramyxovirinae N and P. *J. Gen. Virol.* **84**, 3239–3252.
2. Blanchard, L., Tarbouriech, N., Blackledge, M., Timmins, P., Burmeister, W. P., Ruigrok, R. W. & Marion, D. (2004). Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. *Virology*, **319**, 201–211.
3. Tarbouriech, N., Curran, J., Ruigrok, R. W. & Burmeister, W. P. (2000). Tetrameric coiled coil domain of Sendai virus phosphoprotein. *Nature Struct. Biol.* **7**, 777–781.
4. Llorente, M. T., Garcia-Barreno, B., Calero, M., Camafeita, E., Lopez, J. A., Longhi, S. *et al.* (2006). Structural analysis of the human respiratory syncytial virus phosphoprotein: characterization of an alpha-helical domain involved in oligomerization. *J. Gen. Virol.* **87**, 159–169.
5. Bourhis, J. M., Canard, B. & Longhi, S. (2006). Structural disorder within the replicative complex of measles virus: functional implications. *Virology*, **344**, 94–110.
6. Mavrakis, M., McCarthy, A. A., Roche, S., Blondel, D. & Ruigrok, R. W. (2004). Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. *J. Mol. Biol.* **343**, 819–831.
7. Ding, H., Green, T. J., Lu, S. & Luo, M. (2006). Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J. Virol.* **80**, 2808–2814.
8. Ding, H., Green, T. J. & Luo, M. (2004). Crystallization and preliminary X-ray analysis of a proteinase-K-resistant domain within the phosphoprotein of vesicular stomatitis virus (Indiana). *Acta Crystallogr. D*, **60**, 2087–2090.
9. Mavrakis, M., Mehouas, S., Real, E., Iseni, F., Blondel, D., Tordo, N. & Ruigrok, R. W. (2006). Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. *Virology*, **349**, 422–429.
10. Chen, M., Ogino, T. & Banerjee, A. K. (2007). Interaction of vesicular stomatitis virus P and N proteins: Identification of two overlapping domains at the N-terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA. *J. Virol.* **81**, 13478–13485.
11. Gerard, F., Ribeiro, E., Albertini, A., Zaccai, G., Ebel, C., Ruigrok, R. & Jamin, M. (2007). Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution. *Biochemistry*, **46**, 10328–10338.
12. Jacob, Y., Real, E. & Tordo, N. (2001). Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains. *J. Virol.* **75**, 9613–9622.
13. Gigant, B., Iseni, F., Gaudin, Y., Knossow, M. & Blondel, D. (2000). Neither phosphorylation nor the amino-terminal part of rabies virus phosphoprotein is required for its oligomerization. *J. Gen. Virol.* **81**, 1757–1761.
14. Ribeiro, E. A., Jr, Favier, A., Gerard, F. C., Leyrat, C., Brutscher, B., Blondel, D. *et al.* (2008). Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein. *J. Mol. Biol.* **382**, 525–538.
15. Chenik, M., Chebli, K. & Blondel, D. (1995). Translation initiation at alternate in-frame AUG codons in the rabies virus phosphoprotein mRNA is mediated by a ribosomal leaky scanning mechanism. *J. Virol.* **69**, 707–712.
16. Pasdeloup, D., Poisson, N., Raux, H., Gaudin, Y., Ruigrok, R. W. & Blondel, D. (2005). Nucleocytoplasmic shuttling of the rabies virus P protein requires a nuclear localization signal and a CRM1-dependent nuclear export signal. *Virology*, **334**, 284–293.
17. Moseley, G. W., Filmer, R. P., DeJesus, M. A. & Jans, D. A. (2007). Nucleocytoplasmic distribution of rabies virus P-protein is regulated by phosphorylation adjacent to C-terminal nuclear import and export signals. *Biochemistry*, **46**, 12053–12061.
18. Vidy, A., Bougrini, J., Cheldi-Alix, M. & Blondel, D. (2007). The nulceocytoplasmic rabies P protein counteracts interferons signaling by inhibiting both nulear accumulation and DNA binding of STAT1. *J. Virol.* **81**, 4255–4263.

19. Blondel, D., Regad, T., Poisson, N., Pavie, B., Harper, F., Pandolfi, P. P. *et al.* (2002). Rabies virus P and small P products interact directly with PML and reorganize PML nuclear bodies. *Oncogene*, **21**, 7957–7970.
20. Vidy, A., Chelbi-Alix, M. & Blondel, D. (2005). Rabies virus P protein interacts with STAT1 and inhibits interferon signal transduction pathways. *J. Virol.* **79**, 14411–14420.
21. Raux, H., Flamand, A. & Blondel, D. (2000). Interaction of the rabies virus P protein with the LC8 dynein light chain. *J. Virol.* **74**, 10212–10216.
22. Jacob, Y., Badrane, H., Ceccaldi, P. E. & Tordo, N. (2000). Cytoplasmic dynein LC8 interacts with lyssavirus phosphoprotein. *J. Virol.* **74**, 10217–10222.
23. Perutz, M. F. (1963). X-ray analysis of hemoglobin. *Science*, **140**, 863–869.
24. Blake, C. C., Koenig, D. F., Mair, G. A., North, A. C., Phillips, D. C. & Sarma, V. R. (1965). Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Å resolution. *Nature*, **206**, 757–761.
25. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. & Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, **181**, 662–666.
26. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science*, **181**, 223–230.
27. Tanford, C. (1968). Protein denaturation. *Adv. Protein Chem.* **23**, 121–282.
28. Wright, P. E. & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–331.
29. Uversky, V. N., Gillespie, J. R. & Fink, A. L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins: Struct. Funct. Genet.* **41**, 415–427.
30. Uversky, V. N. (2002). What does it mean to be natively unfolded? *Eur. J. Biochem.* **269**, 2–12.
31. Tompa, P. (2002). Intrinsically unstructured proteins. *Trends Biochem. Sci.* **27**, 527–533.
32. Spolar, R. S. & Record, M. T., Jr (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **263**, 777–784.
33. Dyson, H. J. & Wright, P. E. (2002). Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **12**, 54–60.
34. Sugase, K., Dyson, H. J. & Wright, P. E. (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, **447**, 920–921.
35. Demarest, S. J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H. J. *et al.* (2002). Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature*, **415**, 549–553.
36. Uversky, V. N., Segel, D. J., Doniach, S. & Fink, A. L. (1998). Association-induced folding of globular proteins. *Proc. Natl Acad. Sci. USA*, **95**, 5480–5483.
37. Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N. & Dunker, A. K. (2006). Intrinsic disorder in transcription factors. *Biochemistry*, **45**, 6873–6888.
38. Bourhis, J. M., Johansson, K., Receveur-Brechot, V., Oldfield, C. J., Dunker, K. A., Canard, B. & Longhi, S. (2004). The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner. *Virus Res.* **99**, 157–167.
39. Romero, P., Obradovic, Z. & Dunker, A. K. (2004). Natively disordered proteins: functions and predictions. *Appl. Bioinformatics*, **3**, 105–113.
40. Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z. & Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **32**, 1037–1049.
41. Tompa, P. & Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. *FASEB J.* **18**, 1169–1175.
42. Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F. & Jones, D. T. (2004). The DISOPRED server for the prediction of protein disorder. *Bioinformatics*, **20**, 2138–2139.
43. Bourhis, J. M., Canard, B. & Longhi, S. (2007). Predicting protein disorder and induced folding: from theoretical principles to practical applications. *Curr. Protein Pept. Sci.* **8**, 135–149.
44. Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S. *et al.* (2005). FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*, **21**, 3435–3438.
45. Promponas, V. J., Enright, A. J., Tsoka, S., Kreil, D. P., Leroy, C., Hamodrakas, S. *et al.* (2000). CAST: an iterative algorithm for the complexity analysis of sequence tracts. Complexity analysis of sequence tracts. *Bioinformatics*, **16**, 915–922.
46. Bernado, P., Blanchard, L., Timmins, P., Marion, D., Ruigrok, R. W. & Blackledge, M. (2005). A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc. Natl Acad. Sci. USA*, **102**, 17002–17007.
47. Tarbouriech, N., Curran, J., Ebel, C., Ruigrok, R. W. & Burmeister, W. P. (2000). On the domain structure and the polymerization state of the sendai virus P protein. *Virology*, **266**, 99–109.
48. Houben, K., Blanchard, L., Blackledge, M. & Marion, D. (2007). Intrinsic dynamics of the partly unstructured PX domain from the Sendai virus RNA polymerase co-factor P. *Biophys. J.* **93**, 2930–2940.
49. Lieutaud, P., Canard, B. & Longhi, S. (2008). MeDor: a metaserver for predicting protein disorder. *BMC Genomics*, **9**, S25.
50. Vucetic, S., Brown, C. J., Dunker, A. K. & Obradovic, Z. (2003). Flavors of protein disorder. *Proteins: Struct. Funct. Genet.* **52**, 573–584.
51. Ferron, F., Longhi, S., Canard, B. & Karlin, D. (2006). A practical overview of protein disorder prediction methods. *Proteins: Struct. Funct. Genet.* **65**, 1–14.
52. King, R. D., Ouali, M., Strong, A. T., Aly, A., Elmaghraby, A., Kantardzic, M. & Page, D. (2000). Is it better to combine predictions? *Protein Eng.* **13**, 15–19.
53. Schulz, G. E., Barry, C. D., Friedman, J., Chou, P. Y., Fasman, G. D., Finkelstein, A. V. *et al.* (1974). Comparison of predicted and experimentally determined secondary structure of adenyl kinase. *Nature*, **250**, 140–142.
54. Brown, C. J., Takayama, S., Campen, A. M., Vise, P., Marshall, T. W., Oldfield, C. J. *et al.* (2002). Evolutionary rate heterogeneity in proteins with long disordered regions. *J. Mol. Evol.* **55**, 104–110.
55. Pei, J. & Grishin, N. V. (2001). AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics*, **17**, 700–712.
56. Kingston, R. L., Hamel, D. J., Gay, L. S., Dahlquist, F. W. & Matthews, B. W. (2004). Structural basis for

the attachment of a paramyxoviral polymerase to its template. *Proc. Natl Acad. Sci. USA*, **101**, 8301–8306.
57. Ferron, F., Rancurel, C., Longhi, S., Cambillau, C., Henrissat, B. & Canard, B. (2005). VaZyMolO: a tool to define and classify modularity in viral proteins. *J. Gen. Virol.* **86**, 743–749.
58. Emerson, S. U. & Schubert, M. (1987). Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus. *Proc. Natl Acad. Sci. USA*, **84**, 5655–5659.
59. Green, T. J., Macpherson, S., Qiu, S., Lebowitz, J., Wertz, G. W. & Luo, M. (2000). Study of the assembly of vesicular stomatitis virus N protein: role of the P protein. *J. Virol.* **74**, 9515–9524.
60. Das, T., Pattnaik, A. K., Takacs, A. M., Li, T., Hwang, L. N. & Banerjee, A. K. (1997). Basic amino acid residues at the carboxy-terminal eleven amino acid region of the phosphoprotein (P) are required for transcription but not for replication of vesicular stomatitis virus genome RNA. *Virology*, **238**, 103–114.
61. Takacs, A. M., Barik, S., Das, T. & Banerjee, A. K. (1992). Phosphorylation of specific serine residues within the acidic domain of the phosphoprotein of vesicular stomatitis virus regulates transcription in vitro. *J. Virol.* **66**, 5842–5848.
62. Barik, S. & Banerjee, A. K. (1992). Phosphorylation by cellular casein kinase II is essential for transcriptional activity of vesicular stomatitis virus phosphoprotein P. *Proc. Natl Acad. Sci. USA*, **89**, 6570–6574.
63. Barik, S. & Banerjee, A. K. (1992). Sequential phosphorylation of the phosphoprotein of vesicular stomatitis virus by cellular and viral protein kinases is essential for transcription activation. *J. Virol.* **66**, 1109–1118.
64. Gao, Y., Greenfield, N. J., Cleverley, D. Z. & Lenard, J. (1996). The transcriptional form of the phosphoprotein of vesicular stomatitis virus is a trimer: structure and stability. *Biochemistry*, **35**, 14569–14573.
65. Gao, Y. & Lenard, J. (1995). Cooperative binding of multimeric phosphoprotein (P) of vesicular stomatitis virus to polymerase (L) and template: pathways of assembly. *J. Virol.* **69**, 7718–7723.
66. Gao, Y. & Lenard, J. (1995). Multimerization and transcriptional activation of the phosphoprotein (P) of vesicular stomatitis virus by casein kinase-II. *EMBO J.* **14**, 1240–1247.
67. Chenik, M., Chebli, K., Gaudin, Y. & Blondel, D. (1994). In vivo interaction of rabies virus phosphoprotein (P) and nucleoprotein (N): existence of two N-binding sites on P protein. *J. Gen. Virol.* **75**, 2889–2896.
68. Chenik, M., Schnell, M., Conzelmann, K. K. & Blondel, D. (1998). Mapping the interacting domains between the rabies virus polymerase and phosphoprotein. *J. Virol.* **72**, 1925–1930.
69. Fu, Z. F., Zheng, Y., Wunner, W. W., Koprowski, H. & Dietzschold, B. (1994). Both the N- and C-terminal domains of the nominal phosphoprotein of rabies virus are involved in binding to the nucleoprotein. *Virology*, **200**, 590–597.
70. Scholtz, J. M., Qian, H., York, E. J., Stewart, J. M. & Baldwin, R. L. (1991). Parameters of helix-coil transition theory for alanine-based peptides of varying chain lengths in water. *Biopolymers*, **31**, 1463–1470.
71. Uversky, V. N. (1993). Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry*, **32**, 13288–13298.
72. Albertini, A. A., Clapier, C. R., Wernimont, A. K., Schoehn, G., Weissenhorn, W. & Ruigrok, R. W. (2006). Isolation and crystallization of a unique size category of recombinant Rabies virus Nucleoprotein-RNA rings. *J. Struct. Biol.* **158**, 129–133.
73. Johansson, K., Bourhis, J. M., Campanacci, V., Cambillau, C., Canard, B. & Longhi, S. (2003). Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. *J. Biol. Chem.* **278**, 44567–44573.
74. Curran, J., Boeck, R., Lin-Marq, N., Lupas, A. & Kolakofsky, D. (1995). Paramyxovirus phosphoproteins form homotrimers as determined by an epitope dilution assay, via predicted coiled coils. *Virology*, **214**, 139–149.
75. Curran, J., Marq, J. B. & Kolakofsky, D. (1995). An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. *J. Virol.* **69**, 849–855.
76. Kolakofsky, D., Le Mercier, P., Iseni, F. & Garcin, D. (2004). Viral DNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology*, **318**, 463–473.
77. Das, S. C. & Pattnaik, A. K. (2004). Phosphorylation of vesicular stomatitis virus phosphoprotein P is indispensable for virus growth. *J. Virol.* **78**, 6420–6430.
78. Pattnaik, A. K., Hwang, L., Li, T., Englund, N., Mathur, M., Das, T. & Banerjee, A. K. (1997). Phosphorylation within the amino-terminal acidic domain I of the phosphoprotein of vesicular stomatitis virus is required for transcription but not for replication. *J. Virol.* **71**, 8167–8175.
79. Gupta, A. K., Blondel, D., Choudhary, S. & Banerjee, A. K. (2000). The phosphoprotein of rabies virus is phosphorylated by a unique cellular protein kinase and specific isomers of protein kinase C. *J. Virol.* **74**, 91–98.
80. Byrappa, S., Hendricks, D. D., Pan, Y. B., Seyer, J. M. & Gupta, K. C. (1995). Intracellular phosphorylation of the Sendai virus P protein. *Virology*, **208**, 408–413.
81. Byrappa, S., Pan, Y. B. & Gupta, K. C. (1996). Sendai virus P protein is constitutively phosphorylated at serine249: high phosphorylation potential of the P protein. *Virology*, **216**, 228–234.
82. Huntley, C. C., De, B. P. & Banerjee, A. K. (1997). Phosphorylation of Sendai virus phosphoprotein by cellular protein kinase C zeta. *J. Biol. Chem.* **272**, 16578–16584.
83. Radivojac, P., Obradovic, Z., Smith, D. K., Zhu, G., Vucetic, S., Brown, C. J. *et al.* (2004). Protein flexibility and intrinsic disorder. *Protein Sci.* **13**, 71–80.
84. Lenard, J. (1999). Host cell protein kinases in nonsegmented negative-strand virus (mononegavirales) infection. *Pharmacol. Ther.* **83**, 39–48.
85. Chattopadhyay, D., Raha, T. & Chattopadhyay, D. (1997). Single serine phosphorylation within the acidic domain of Chandipura virus P protein regulates the transcription in vitro. *Virology*, **239**, 11–19.
86. Raha, T., Samal, E., Majumdar, A., Basak, S., Chattopadhyay, D. & Chattopadhyay, D. J. (2000). N-terminal region of P protein of Chandipura virus is responsible for phosphorylation-mediated homodimerization. *Protein Eng.* **13**, 437–444.

87. Masters, P. S. & Banerjee, A. K. (1988). Resolution of multiple complexes of phosphoprotein NS with nucleocapsid protein N of vesicular stomatitis virus. *J. Virol.* **62**, 2651–2657.
88. Takacs, A. M., Das, T. & Banerjee, A. K. (1993). Mapping of interacting domains between the nucleocapsid protein and the phosphoprotein of vesicular stomatitis virus by using a two-hybrid system. *Proc. Natl Acad. Sci. USA*, **90**, 10375–10379.
89. Tompa, P. (2005). The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.* **579**, 3346–3354.
90. Bourhis, J. M., Receveur-Brechot, V., Oglesbee, M., Zhang, X., Buccellato, M., Darbon, H. *et al.* (2005). The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded. *Protein Sci.* **14**, 1975–1992.
91. Li, X., Romero, P., Rani, M., Dunker, A. K. & Obradovic, Z. (1999). Predicting protein disorder for N-, C-, and internal regions. *Genome Inform. Ser. Workshop Genome Inform.* **10**, 30–40.
92. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J. & Dunker, A. K. (2001). Sequence complexity of disordered protein. *Proteins: Struct. Funct. Genet.* **42**, 38–48.
93. Liu, J., Tan, H. & Rost, B. (2002). Loopy proteins appear conserved in evolution. *J. Mol. Biol.* **322**, 53–64.
94. Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, **21**, 3433–3434.
95. Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J. & Russell, R. B. (2003). Protein disorder prediction: implications for structural proteomics. *Structure*, **11**, 1453–1459.
96. Yang, Z. R., Thomson, R., McNeil, P. & Esnouf, R. M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, **21**, 3369–3376.
97. Coeytaux, K. & Poupon, A. (2005). Prediction of unfolded segments in a protein sequence based on amino acid composition. *Bioinformatics*, **21**, 1891–1900.
98. Wooton, J. (1994). Non-globular doamins in protein sequences: automated segmentation using complexity measures. *Comput. Chem.* **18**, 269–285.
99. Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J. *et al.* (1997). Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell Mol. Life Sci.* **53**, 621–645.
100. Combet, C., Blanchet, C., Geourjon, C. & Deleage, G. (2000). NPS@: network protein sequence analysis. *Trends Biochem. Sci.* **25**, 147–150.
101. King, R. D., Saqi, M., Sayle, R. & Sternberg, M. J. (1997). DSC: public domain protein secondary structure predication. *Comput. Appl. Biosci.* **13**, 473–474.
102. Garnier, J., Gibrat, J. F. & Robson, B. (1996). GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol.* **266**, 540–553.
103. Guermeur, Y., Geourjon, C., Gallinari, P. & Deleage, G. (1999). Improved performance in protein secondary structure prediction by inhomogeneous score combination. *Bioinformatics*, **15**, 413–421.
104. Rost, B., Sander, C. & Schneider, R. (1994). PHD – an automatic mail server for protein secondary structure prediction. *Comput. Appl. Biosci.* **10**, 53–60.
105. Frishman, D. & Argos, P. (1997). Seventy-five percent accuracy in protein secondary structure prediction. *Proteins: Struct. Funct. Genet.* **27**, 329–335.
106. Levin, J. M. (1997). Exploring the limits of nearest neighbour secondary structure prediction. *Protein Eng.* **10**, 771–776.
107. Kneller, D. G., Cohen, F. E. & Langridge, R. (1990). Improvements in protein secondary structure prediction by an enhanced neural network. *J. Mol. Biol.* **214**, 171–182.
108. Pollastri, G. & McLysaght, A. (2005). Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics*, **21**, 1719–1720.
109. Rost, B. & Liu, J. (2003). The PredictProtein server. *Nucleic Acids Res.* **31**, 3300–3304.
110. Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202.
111. Pollastri, G., Przybylski, D., Rost, B. & Baldi, P. (2002). Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins: Struct. Funct. Genet.* **47**, 228–235.
112. Ouali, M. & King, R. D. (2000). Cascaded multiple classifiers for secondary structure prediction. *Protein Sci.* **9**, 1162–1176.
113. Lupas, A., Van Dyke, M. & Stock, J. (1991). Predicting coiled coils from protein sequences. *Science*, **252**, 1162–1164.
114. Berger, B., Wilson, D. B., Wolf, E., Tonchev, T., Milla, M. & Kim, P. S. (1995). Predicting coiled coils by use of pairwise residue correlations. *Proc. Natl Acad. Sci. USA*, **92**, 8259–8263.
115. Wolf, E., Kim, P. S. & Berger, B. (1997). MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci.* **6**, 1179–1189.
116. Delorenzi, M. & Speed, T. (2002). An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics*, **18**, 617–625.
117. Notredame, C., Higgins, D. G. & Heringa, J. (2000). T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217.
118. Guarente, L. (1993). Strategies for the identification of interacting proteins. *Proc. Natl Acad. Sci. USA*, **90**, 1639–1641.
119. Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry*, **6**, 1948–1954.
120. Luo, P. & Baldwin, R. (1997). Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry*, **36**, 8413–8421.

III. Conclusion

Les résultats présentés dans cet article soulignent l'organisation modulaire des dimères de phosphoprotéine avec la présence de domaines capables de se replier indépendamment du reste de la protéine, et séparés par de longs segments désordonnés. L'organisation modulaire est remarquablement conservée entre les genres *Lyssavirus* et *Vesiculovirus*, mais également entre les familles *Rhabdoviridae* et *Paramyxoviridae,* soulignant ainsi l'importance fonctionnelle de la présence d'une région N-terminale désordonnée se liant à N^0 (Mavrakis, Mehouas et al. 2006; Chen, Ogino et al. 2007), d'un domaine central d'oligomérisation (Ding, Green et al. 2006) et d'un module C-terminal responsable de l'interaction avec les nucléocapsides virales (Jacob, Real et al. 2001; Schoehn, Iseni et al. 2001; Mavrakis, McCarthy et al. 2004).

Il émerge de cette étude que la région fonctionnelle la moins bien caractérisée structuralement est la région N-terminale de la phosphoprotéine, impliquée dans l'interaction avec N^0 ainsi qu'avec la protéine L (Castel, Chteoui et al. 2009). Ce constat nous a conduits à entreprendre une caractérisation biophysique et structurale plus poussée des fragments N-terminaux intrinsèquement désordonnés des phosphoprotéines de VSV et de RV, en utilisant notamment la diffusion de rayons X aux petits angles (SAXS) couplée à une analyse par optimisation d'ensembles de structures atomiques désordonnées (EOM) (Bernado, Blanchard et al. 2005; Bernado, Mylonas et al. 2007), la résonance magnétique nucléaire (RMN) et la spectroscopie de dichroïsme circulaire (CD).

CHAPITRE III : ETUDES BIOPHYSIQUES ET STRUCTURALES DE LA REGION N-TERMINALE DE LA PHOSPHOPROTEINE DES RHABDOVIRIDAE

« Si tu veux boire du lait, restes chez toi ! »

(Euripedes de Almeida Ribeiro Jr, Drôme Provençale, Juin 2010)

I. Introduction

Le travail présenté dans ce chapitre porte sur la caractérisation biophysique et structurale de la région N-terminale de la phosphoprotéine du VSV et du virus de la rage, comprenant des travaux de RMN réalisés en collaboration avec Malene Ringkjøbing Jensen et Martin Blackledge du laboratoire Protein Dynamics and Flexibility à l'Institut de Biologie Structurale. En combinant les techniques de RMN, de SAXS et de CD, nous avons pu montrer que la région N-terminale (résidus 1-60) de P chez le VSV (VSV P_{NTD}) existe en solution sous la forme d'un ensemble structural peuplé de conformères de compacité différente et présente des hélices α transitoires en proportions significatives dans deux régions de la molécule. Les populations de conformères étendus ou compacts sont affectées par la présence d'agents stabilisants ou dénaturants, ce qui se traduit également par des modifications dans le spectre de dichroïsme circulaire. Par ailleurs, le calcul du pourcentage d'hélices α obtenu à partir des structures sélectionnées sur base des données de SAXS en utilisant l'approche E.O.M. (optimisation d'un ensemble de modèles par un algorithme génétique contre un profile SAXS expérimental) montre une bonne corrélation avec les données de RMN et de CD.

Nous avons également caractérisé le désordre structural de la région N-terminale de P du virus de la rage (résidus 1-68) (RV P_{NTD}), en se focalisant cette fois sur l'utilisation du SAXS et de l'optimisation d'ensembles pour étudier notamment l'effet de la température sur la structure de RV P_{NTD}. En particulier, nous nous sommes intéressés à l'influence de la température, et d'agents stabilisants ou dénaturants sur le contenu en hélices α et en polyproline II.

L'étude par SAXS montre que le peptide de RV se compacte et se replie à haute concentration, un phénomène qui est beaucoup moins marqué avec celui de VSV, et qui est probablement lié à des répulsions plus fortes au sein de la molécule. La région N-terminale de la phosphoprotéine du virus de la rage est sensiblement plus chargée que celle de VSV. L'effet du repliement induit par l'addition de 1 M de triméthylamine N-oxyde (TMAO) est moins marqué chez RV que chez VSV, une observation qui est cohérente avec la plus faible propension intrinsèque à former des hélices α observée pour la région P_{NTD} du RV que pour celle du VSV.

Nous nous sommes aussi intéressés à l'effet de la température sur la structuration du RV P_{NTD}. Ce phénomène a été étudié par spectroscopie de dichroïsme circulaire par diverses équipes de recherche. L'origine physique du changement structural observé fait l'objet d'un débat considérable dans la littérature. En particulier, différents auteurs l'ont attribué soit à une perte de conformations en polyproline II, soit à une augmentation du pourcentage d'hélices α (Uversky, Li et al. 2001; Dawson, Muller et al. 2003; Gast, Zirwer et al. 2003; Jarvet, Damberg et al. 2003; Sanchez-Puig, Veprintsev et al. 2005; Malm, Jonsson et al. 2007; Jeganathan, von Bergen et al. 2008; Uversky 2009; Kjaergaard, Norholm et al. 2010).

II. ARTICLE II: The N^0 Binding region of the Vesicular Stomatitis Virus Phosphoprotein is Globally Disordered but Contains Transient α-Helices

Auteurs: Cédric Leyrat, Malene Ringkjøbing Jensen, Euripedes de Almeida Ribeiro Jr, Francine Gérard, Rob Ruigrok, Martin Blackledge et Marc Jamin.

Soumis à *Protein Science*

THE N^0-BINDING REGION OF THE VESICULAR STOMATITIS VIRUS PHOSPHOPROTEIN IS GLOBALLY DISORDERED BUT CONTAINS TRANSIENT α-HELICES

Cédric Leyrat[1], Malene Ringkjøbing Jensen[2], Euripedes de Almeida Ribeiro Jr[1], Francine Gérard[1], Rob W. H. Ruigrok[1], Martin Blackledge[2] and Marc Jamin[1*]

[1]UMI 3265 UJF-EMBL-CNRS, Unit of Virus Host Cell Interactions
6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France
[2]Protein Dynamics and Flexibility, Institut de Biologie Structurale, UMR 5075 CEA-CNRS-UJF, 41 rue Jules Horowitz, Grenoble 38027, France

* **Corresponding author:** Unit of Virus Host Cell Interactions (UVHCI), UMI 3265 UJF-EMBL-CNRS, 6, rue Jules Horowitz, B.P. 181, 38042 Grenoble Cedex 9, France
E-mail : jamin@embl.fr, Phone: + 33 4 76 20 94 62, Fax : + 33 4 76 20 94 00

ABSTRACT

The phosphoprotein (P) of vesicular stomatitis virus (VSV) interacts with nascent nucleoprotein (N), forming the N^0-P complex that is indispensable for the correct encapsidation of newly synthesized viral RNA genome. In this complex, the N-terminal region (P_{NTD}) of P prevents N from binding to cellular RNA and keeps it available for encapsidating viral RNA genomes. Here, using nuclear magnetic resonance spectroscopy and small angle X-ray scattering, we show that isolated P_{NTD}, corresponding to the 60 first N-terminal residues of VSV P, has overall molecular dimensions and a dynamic behavior characteristic of a disordered protein but transiently populates conformers containing α-helices. The modeling of P_{NTD} as a conformational ensemble by the ensemble optimization method (EOM) using SAXS data correctly reproduces the α-helical content detected by NMR spectroscopy and suggests the co-existence of sub-ensembles of different compactness. The populations and overall dimensions of these sub-ensembles are affected by the addition of stabilizing (1 M TMAO) or destabilizing (6 M GdmCl) co-solvents. Our results are interpreted in the context of a scenario whereby VSV P_{NTD} constitutes a molecular recognition element (MoRE) undergoing a disorder-to-order transition upon binding to its partner when forming the N^0-P complex.

KEYWORDS: Rhabdovirus, vesicular stomatitis virus, phosphoprotein, SAXS, NMR, intrinsically disordered proteins, molecular recognition element

SHORT STATEMENT: The complex between the N-terminal part of the phosphoprotein (P_{NTD}) and nascent nucleoprotein (N^0) is essential for the replication of the vesicular stomatitis virus (VSV). Here, using SAXS and NMR spectroscopy, we showed that a peptide of 60 amino acids encompassing the N^0-binding site is globally disordered but contains transient α-helices suggesting that P_{NTD} is a molecular recognition element (MoRE) that undergoes a disorder-to-order transition upon binding to its partner.

INTRODUCTION

In the last decade, our understanding of the relationships between protein structure and function was revolutionized by the discoveries that numerous proteins are intrinsically disordered

(IDPs) or contain intrinsically disordered regions (IDRs) and that these disordered proteins are implicated in various biological functions (Wright and Dyson 1999; Uversky, Gillespie et al. 2000; Dunker, Brown et al. 2002; Tompa 2002; Wright and Dyson 2009). These findings continuously raise new challenges in structural biology for predicting IDPs or IDRs from the amino acid sequence and for characterizing their structural properties. NMR spectroscopy (Dyson and Wright 2004; Eliezer 2009; Jensen, Markwick et al. 2009) and small-angle scattering (Bernado, Mylonas et al. 2007; Bernado and Blackledge 2009) revealed to be particularly suitable for providing detailed structural information on these proteins. Various algorithms were developed for identifying IDPs or for locating IDRs (Gérard, Ribeiro et al. 2009; He, Wang et al. 2009). Complete genome surveys revealed that not all kingdoms of life are equal when we consider the abundance of IDPs and IDRs encoded in the genomes. A large fraction (20 – 40 %) of eukaryote genomes encodes IDP or proteins containing long IDRs, whereas they are much less abundant (< 10%) in archea and bacteria (Dunker, Obradovic et al. 2000; Ward, Sodhi et al. 2004; Oldfield, Cheng et al. 2005). Viruses, and in particular RNA viruses, occupy an intermediate position between eukaryotes and bacteria, and exhibit a high rate of loosely packed proteins (Tokuriki, Oldfield et al. 2009).

The *Rhabdoviridae* is a family of non-segmented negative-sense RNA viruses, which includes human (rabies virus - RV), animal (vesicular stomatitis virus - VSV) and plant pathogens. The machinery of RNA transcription and replication of these viruses is composed of the genomic RNA and of three viral proteins: the nucleoprotein (N), the phosphoprotein (P) and the large subunit of the RNA-dependent RNA polymerase (L). The RNA genome is encapsidated by N. Every N binds to nine nucleotides, forming a long helical N-RNA complex (Fields, Knipe et al. 1996) that serves as a template for both transcription and replication (Arnheiter, Davis et al. 1985). The crystal structure of circular N-RNA complexes of both VSV and RV revealed that the N protein contains a N-terminal domain (NTD) and a C-terminal (CTD) domain separated by a positively charged groove which binds the RNA (Albertini, Wernimont et al. 2006; Green, Zhang et al. 2006). NTD and CTD act like jaws that close around and completely enwrap the RNA, forming multiple salt bridges with the phosphate backbone of the RNA. The phosphoprotein (P) is involved in different stages of the viral replication process (Emerson and Yu 1975; Masters and Banerjee 1988). Firstly, it associates with the L protein, forming active RNA polymerase complexes that catalyze both transcription and replication of the viral genome. In this complexes, L carries out the RNA synthesis and the 5' and 3' RNA processing required for the production of messenger RNAs, whereas P acts as an essential cofactor (Emerson and Yu 1975) by attaching L to its N-RNA template (Mellon and Emerson 1978; Emerson and Schubert 1987). P possesses binding sites for both L and the N-RNA complex and thus ensures the correct positioning of the polymerase onto its template. Secondly, P is essential for encapsidating newly synthesized RNA genomes (Peluso 1988; Peluso and Moyer 1988; Howard and Wertz 1989). P binds to nascent N molecules, forming a N^0-P complex (the superscript ^{0} denotes the absence of RNA), which prevents N from binding to cellular RNAs and maintains N in a soluble form, available for the encapsidation of newly synthesized RNA genomes (Masters and Banerjee 1988; Peluso and Moyer 1988; Howard and Wertz 1989; Mavrakis, Iseni et al. 2003; Mavrakis, Mehouas et al. 2006; Chen, Ogino et al. 2007). In both VSV and RV P, the N-terminal region of the protein (P_{NTD}), residues 4 to 40 in RV and residues 11 to 30 in VSV, is sufficient for maintaining N in a soluble and monomeric form (Mavrakis, Mehouas et al. 2006; Chen, Ogino et al. 2007). Because P_{NTD} is rich in negatively charged residues, it was hypothesized that P_{NTD} binds in the RNA-binding cavity of N (Hudson, Condra et al. 1986; Mavrakis, Mehouas et al. 2006), thereby preventing the non-specific attachment of N^0 to cellular RNA.

Recently, we showed that both RV and VSV P form dimers in solution (Gérard, Ribeiro et al. 2007) and are modular proteins made of structured domains and intrinsically disordered regions (IDRs) (Gérard, Ribeiro et al. 2009). The consensus prediction of disordered regions in both RV and VSV P obtained by combining results from different disorder predictors suggested that the 39 N-terminal amino acids of VSV P and the 52 N-terminal amino acids of RV P are structured (Ribeiro, Favier et al. 2008; Gérard, Ribeiro et al. 2009), whereas the following residues (aa 40 to 90 in VSV; aa 53-86 in RV) would be disordered. However, in a previous study, an isolated peptide corresponding to the N-terminal region of RV P (RV P_{NTD}; aa 1-68) appeared to be disordered in

solution (Gérard, Ribeiro et al. 2009). Static light scattering revealed that RV P_{NTD} was monomeric, its Stokes' radius measured by size exclusion chromatography (SEC) corresponded to that expected for an unfolded protein of the same molecular mass and its mean secondary structure content measured by circular dichroism was less than 5% (Gérard, Ribeiro et al. 2009). These contradicting features, namely predictions of a folded domain in P_{NTD} and experimental results showing that P_{NTD} is disordered in solution, are reminiscent of some recently identified short protein motifs that fold only upon binding to their partners and were named molecular recognition elements (MoREs) or molecular recognition features (MoRFs) (Dyson and Wright 2005; Mohan, Oldfield et al. 2006; Sugase, Dyson et al. 2007; Jensen, Houben et al. 2008).

To further investigate the structural properties of the N-terminal region of P, we studied a peptide corresponding to the 60 N-terminal residues of VSV P (P_{NTD}) by nuclear magnetic resonance (NMR) spectroscopy and small angle X-ray scattering (SAXS). Although the NMR spectra exhibit typical features of disordered proteins, transiently formed secondary structures were identified from NMR chemical shifts and the dynamics of the polypeptide chain was probed by ^{15}N relaxation rates. In addition, the average compactness of P_{NTD} was probed by SAXS, and a conformational ensemble that accounts for all experimental observations was built by modeling SAXS data with the Ensemble Optimization Method (EOM). Also, the effects of denaturing or stabilizing co-solvents on the size distribution of the representative ensemble were measured by SAXS. Our results revealed that, in solution, P_{NTD} is a highly dynamic protein transiently forming α-helical elements and populating compact conformers that could play a major role in the formation of the N^0-P complex.

RESULTS

Isolated VSV P_{NTD} is intrinsically disordered but exhibits transient secondary structures

The N-terminal domain of VSV P (P_{NTD}, aa 1-60) was produced in *Escherichia coli* as a recombinant protein containing a C-terminal His-tag (8 aa). The long term stability of P_{NTD} was significantly improved in the presence of a mixture of 50 mM L-Arg and 50 mM L-Glu (Golovanov, Hautbergue et al. 2004).

Under these conditions, an average molecular mass of 7.2 ± 0.7 kDa was determined by size exclusion chromatography combined with detection by multi-angle laser light scattering and refractometry (SEC/MALLS/RI), indicating that P_{NTD} was monomeric (MM calculated from the sequence = 8,054 Da) and well-behaved in solution up to concentrations of 2 mM (Figure 1). Its Stokes' radius of 2.3 ± 0.1 nm indicated an extended conformation, closer to that expected for an unfolded protein of the same molecular mass (R_S = 2.5 nm) than to that expected for a globular protein (R_S = 1.5 nm) (Uversky 1993).

SAXS profiles of P_{NTD} were collected for scattering vector values ($Q = 4\pi(\sin\theta)/\lambda$) ranging from 0.1 to 3.0 nm^{-1} over a concentration range of 5 to 18 $mg.mL^{-1}$. The profiles obtained at different protein concentrations had the same shape and were flat at low Q values indicating the absence of significant aggregation (Figure 2a).

	Protein concentration (mg/mL)	R_g (Guinier) (nm)	R_g (P(r)) (nm)	D_{max} (nm)
50 mM Arg/ 50 mM Glu	5	2.4 ± 0.1	2.4 ± 0.1	7.4
	8	2.4 ± 0.1	2.4 ± 0.1	7.4
	18	2.4 ± 0.1	2.4 ± 0.1	7.4
6 M GdmCl	5	2.9 ± 0.1	2.8 ± 0.1	8.2
1 M TMAO	5	2.3 ± 0.1	2.3 ± 0.1	7.0

Table 1 – Molecular dimensions calculated from SAXS data. Rg values were determined from the Guinier plot or from the distance distribution function (P(r) *versus* r). D_{max} values beyond which P(r) = 0 were adjusted in order that the R_g value obtained with the program GNOM agree with those obtained from the Guinier analysis.

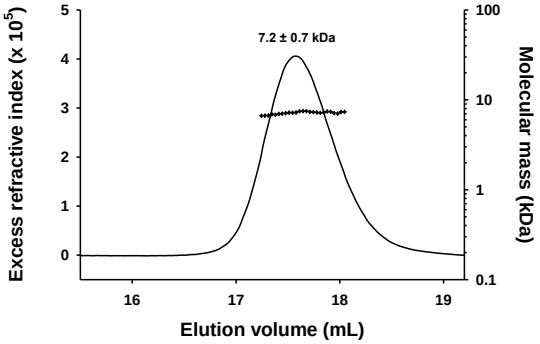


Figure 1. Molecular mass and Stokes' radius of isolated VSV P_{NTD} measured by SEC/MALLS/RI. The line shows the SEC elution profile as monitored by refractometry (left axis). The crosses show the molecular mass calculated from light scattering and refractometry data across the elution peak (right axis). The Stokes's radius (R_S) was determined from the elution volume by calibrating the system with standard proteins of known Stokes'radius (see Material and Methods).

The radius of gyration, R_g, values derived from the Guinier approximation at low Q value (Q.Rg < 1.4) (Figure 2b) or from the distance distribution function (P(r)) (Figure 2c) showed no significant dependence on protein concentration, confirming the absence of aggregation (Table 1). The average R_g value of 2.4 ± 0.1 nm was close to the value of 2.5 nm calculated for a random coil of 68 amino acids (peptide + His-Tag) according to Flory's power-law dependence on chain length (Eq. 2) and using parameters predicted for a random coil with excluded volume (Fitzkee and Rose 2004) and to the value of 2.3 nm calculated from the same power-law dependence using parameters derived from experimental measurements of R_g for IDPs (Bernado and Blackledge 2009). The forward scattering intensity, *I(0)*, was proportional to protein concentration and yielded an estimated molecular mass of 11 ± 1 kDa, in agreement with SEC/MALLS/RI measurements, confirming that P_{NTD} was monomeric and well-behaved in solution. The 2D ^{1}H-^{15}N-HSQC NMR spectrum of P_{NTD} exhibited narrow line widths and poor dispersion of the amide proton chemical shifts (Figure 3), in agreement with structural disorder in solution. These results indicate that VSV P_{NTD} has overall molecular dimensions and NMR spectroscopic properties of a disordered protein, as shown previously for RV P_{NTD} (Gérard, Ribeiro et al. 2009).

The far-UV circular dichroism spectrum exhibited a prominent shoulder near 222 nm and a minimum near 205 nm suggesting the presence of α-helical structure (Figure 4). Assuming that the molar ellipticity value at 222 nm reports only on the presence of α-helix, the average helical content of P_{NTD} was estimated to be 7 % (Luo and Baldwin 1997; Gérard, Ribeiro et al. 2009).

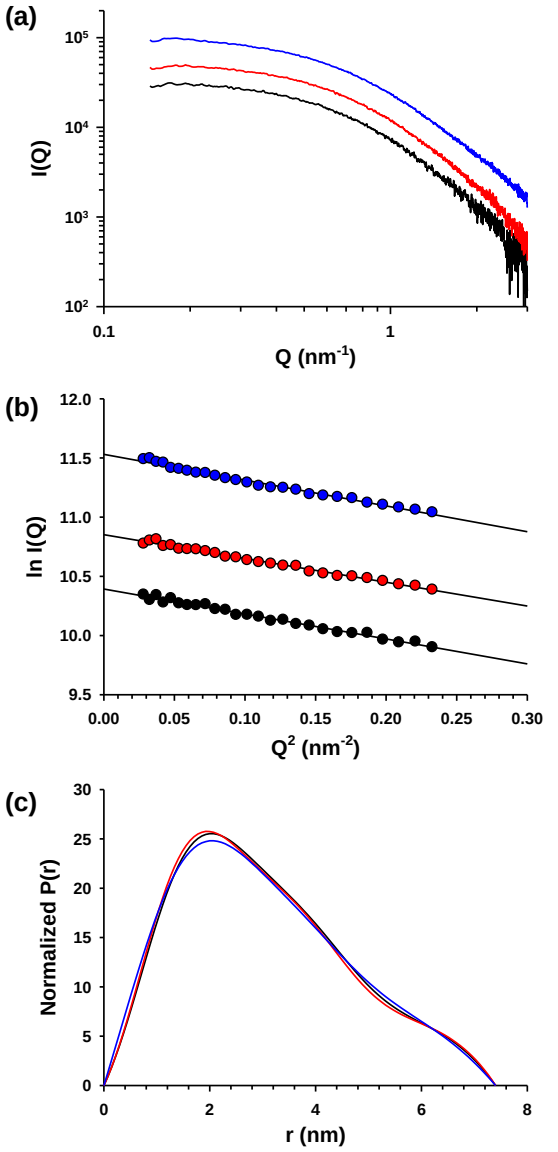


Figure 2. Small-angle X-ray scattering experiments – (a) Scattering curve of P_{NTD} recorded at different concentrations. In this panel and in the following panels b and c protein concentration was at 5 mg.mL^{-1} (black curve or circles), at 8 mg.mL^{-1} (red curve or circles) or at 18 mg.mL^{-1} (blue curve or circles). **(b) Guinier plot.** The fitted lines using 0.10 < Q < 0.28 nm^{-1} yield the radii of gyration shown in Table 1. The fit corresponds to a range of 0.5 < Q.Rg < 1.5. The I(0) value calculated from the intercept, together with protein concentrations yielded a molecular mass estimate of 11 ± 1 kDa. **(c) Distance distribution function.** The distance distribution functions were calculated with GNOM by using D_{max} values shown in Table 1. The surface areas under the curves were normalized to account for the differences in protein concentration.

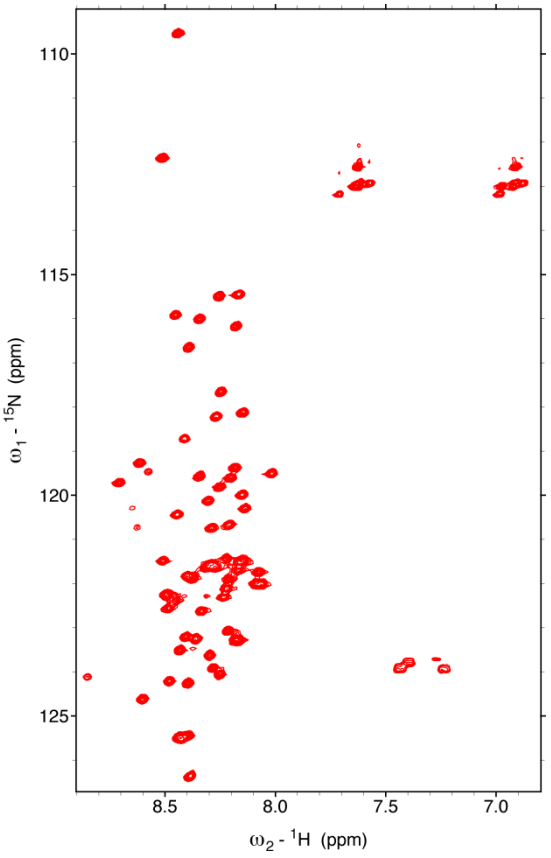


Figure 3. 2D [^{1}H-^{15}N]-HSQC NMR spectrum of VSV P_{NTD}. The spectrum was recorded in 20 mM Tris pH 6.0, 150 mM NaCl, 50 mM Arg, 50 mM Glu at 10 °C.

NMR spectroscopy is highly sensitive to local structure and therefore provides information about the location and population of fluctuating secondary structures within an otherwise flexible protein (Spera and Bax 1991; Wishart, Sykes et al. 1991; Eliezer, Chung et al. 2000; Marsh, Singh et al. 2006; Jensen, Markwick et al. 2009; Jensen, Salmon et al. 2010). After complete assignment of the backbone nuclei (C_α, C_β, C', N and HN), the chemical shifts and ^{15}N nuclear relaxation rates were used to identify transient structural features within P_{NTD}. The secondary structure propensity (SSP) calculated by combining secondary chemical shifts of C_α and C_β showed continuous stretches of positive propensity (SSP > 0.1) for residues 2-12 (helix 1) and 25-38 (helix 2), with the highest population in the region 25-31, suggesting the presence of residual helical structures in these two regions of the protein (Figure 5a) (Marsh, Singh et al. 2006). It also showed a stretch of negative propensity for residues 53-57 characteristic of an extended β structure. Averaging these propensities over P_{NTD} yielded α-helical and β-structure contents of 9 % and 1 %, respectively, in good agreement with the helical content estimated from the CD spectrum. The

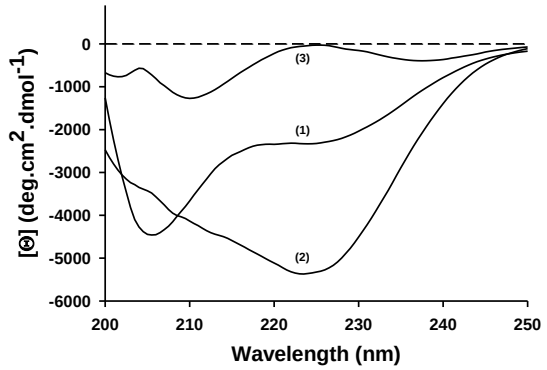


Figure 4. Circular dichroism spectroscopy VSV P_{NTD}. CD spectra of were recorded at 20 °C in 20 mM Tri-HCl pH 7.5 containing 50 mM Arg and 50 mM Glu (1), 1 M TMAO (2), or 6 M GdmCl (3).

uniformly low values of the heteronuclear [^{1}H]-^{15}N NOEs indicated high backbone flexibility on the picosecond to nanosecond time scale (Figure 5b). The presence of transient secondary structures is, however, suggested from the ^{15}N nuclear relaxation rates, R_1 and R_2 (Figure 5c and 5d). In particular, the ^{15}N R_2 rates are larger in the regions of the peptide that have helical propensity according to the SSP score, and the rates oscillate with a periodicity of three-four residues possibly showing the possible effect of a preferential orientation of the diffusion tensor within the helical elements. The regions 3-14 and 26-34 of P_{NTD} exhibited significant intrinsic helical propensity (Figure 5e) (Munoz and Serrano 1994; Lacroix, Viguera et al. 1998) and amino acids that can accept hydrogen bonds from free backbone amide groups and have high N-capping preferences (Doig and Baldwin 1995) were found at the N-terminal extremity of each helical segment. The program AGADIR predicted intrinsic N-capping propensities near 5 % for Asp2 and Asn3 (N_{cap} for helix 1) and of 13.5 % for Asp25 (N_{cap} for helix 2) (Munoz and Serrano 1994; Lacroix, Viguera et al. 1998). These results suggest that the fluctuating helical elements are locally encoded in the amino acid sequence and stabilized by short range interactions.

Modeling the conformational ensemble of VSV P_{NTD}

Recently, different methods were developed for reconstructing low-resolution electron density maps from SAXS data (Svergun, Petoukhov et al. 2001;

Konarev, Petoukhov et al. 2006; Putnam, Hammel et al. 2007; Franke and Svergun 2009). Many of these *ab initio* methods are not suitable for modeling highly flexible molecules because they yield a unique

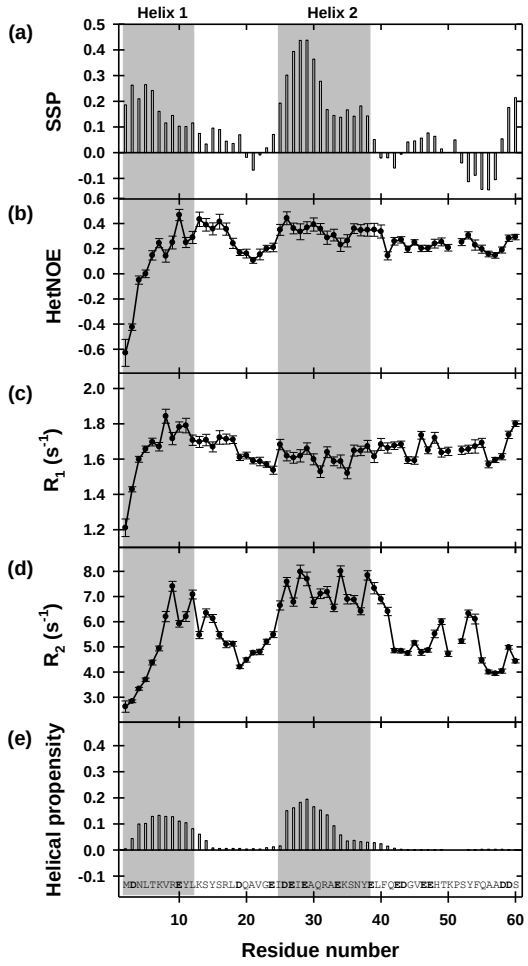


Figure 5. NMR dynamics of isolated VSV P_{NTD}. (a) Secondary structure propensity (SSP). The SSP parameter was calculated from C_α and C_β secondary chemical shifts. **(b) {^{1}H}–^{15}N heteronuclear NOEs. (c,d) Relaxation rates of longitudinal, R_1, and transverse, R_2, magnetization of backbone ^{15}N nuclei.** The shaded area highlight the regions predicted to be helical from the SSP parameter (aa 2-12 and 24-39). NMR data were recorded at 600 MHz and 10 °C using a peptide sample in 20 mM Tris pH 6.0, 150 mM NaCl, 50 mM Arg, 50 mM Glu. **(e) Intrinsic helical propensity.** Bar graph showing the intrinsic helical propensity *per* residue calculated with the program AGADIR (Munoz and Serrano 1994).

structure that at best would provide an average model of the most frequent conformers present in solution. Because VSV P_{NTD} appeared as a disordered protein with fluctuating helical structures, we used the Ensemble Optimization Method (EOM) (Bernado, Mylonas et al. 2007) (Figure 6). This method starts from a large ensemble of randomly generated conformers and selects a sub-ensemble of conformers that collectively reproduces the experimental SAXS data and represents the distribution of structures adopted by the protein in solution (Bernado, Mylonas et al. 2007). Based on the location of residual helical elements identified by NMR spectroscopy, an initial ensemble of 4,500 conformers of P_{NTD} was built using the Flexible Meccano algorithm (Bernado, Blanchard et al. 2005) by pooling an ensemble of 500 conformers containing no helical structure with eight ensembles, of 500 conformers each, containing helices of different lengths (helix 1: aa 2-9 or 2-19, helix 2: aa 25-32 or 25-38) as described in Materials and Methods. The distributions of R_g and D_{max} (the value of maximum dimension) of the initial ensemble were broad and slightly asymmetrical, with R_g values extended from 1.4 to 4.0 nm with a maximum frequency near 2.3 nm (Figure 6a) and with D_{max} values extended from 4.0 to 12.0 nm with a maximum frequency near 7.3 nm, respectively (Figure 6b). The back-calculated SAXS curve from the ensemble of twenty conformers selected by the program EOM reproduced the experimental curve (χ = 0.630) (Supplementary material, Figure S1a). The R_g and D_{max} distributions of the EOM-selected sub-ensemble had widths similar to those of the initial ensemble as expected for a disordered P_{NTD} (Figures 6a and 6b). Figure 6g shows members of the selected ensemble exhibiting varying Rg values as indicated in Figure 6a. However, the R_g distribution within the selected ensemble exhibited two maxima suggesting the presence of two distinct and almost equally important sub-populations of conformers with average R_g values centered on 2.0 and 2.9 nm (Figure 6a). The D_{max} distribution is not as clearly bimodal (Figure 6b). Variation in the number of conformers selected by the EOM procedure from 50 to 2 had no significant effect on the quality of the fit and yielded similar bimodal distributions of R_g values (data not shown). The presence of helical elements in the models introduced specific features in the scattering curve at Q values above 1 nm^{-1} (Supplementary material, Figure S1b), and therefore the average helix content of the selected ensemble of about 7 % was in good agreement with estimations obtained from CD and NMR. Models containing helical elements were found in both sub-populations of compact and extended conformers (Figures 6a and 6g). Successive and independent selections by EOM yielded similar results and similar ensembles of conformers showing that the bimodal distribution

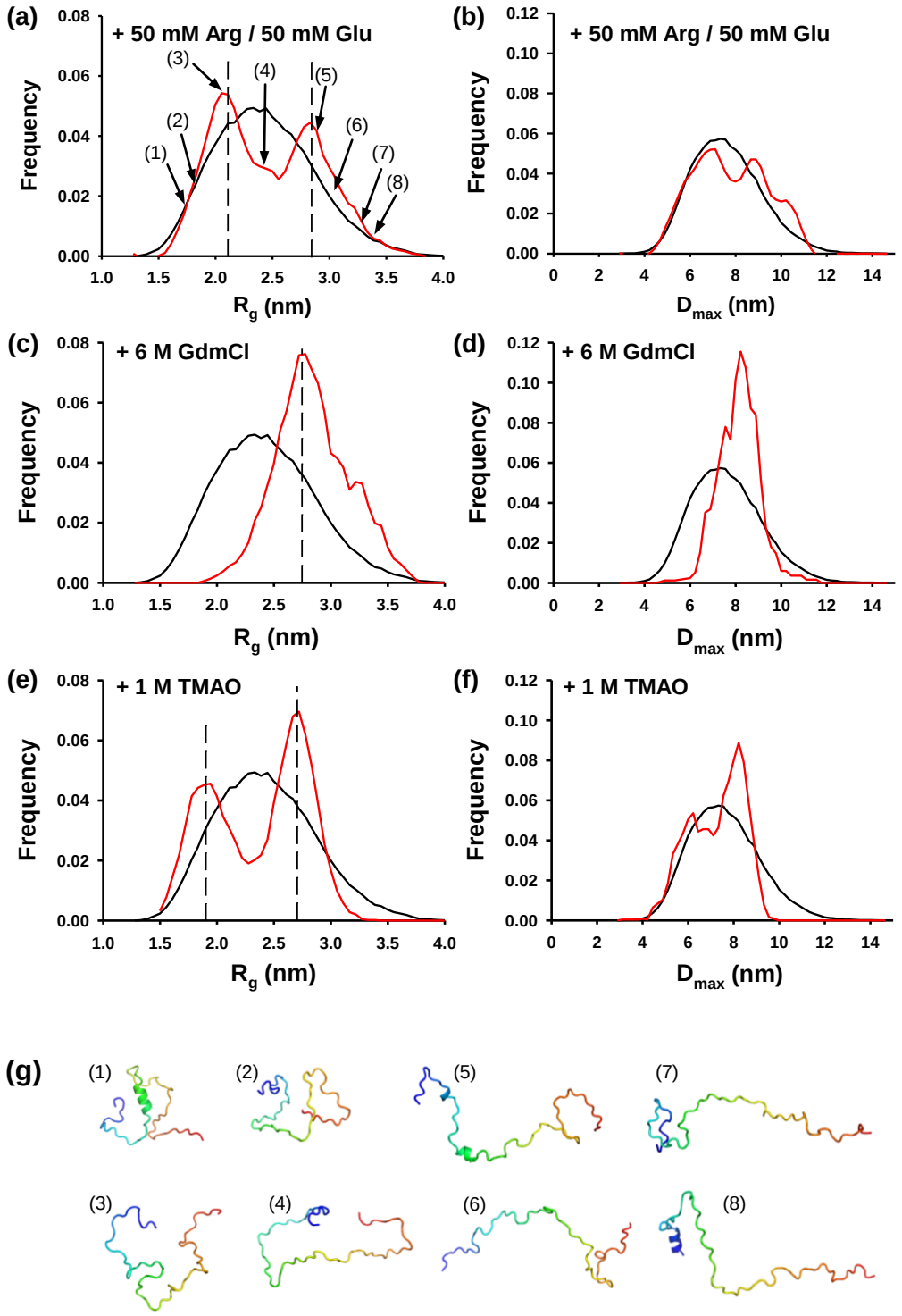


Figure 6. Conformational ensemble selection and effects of stabilizing and destabilizing co-solutes. Ensembles of 20 conformers that collectively reproduce the experimental curves were selected from the initial ensemble. In each figure, the black curve shows the R_g and D_{max} distributions calculated for the initial ensemble of conformers, whereas the red curve shows the R_g and D_{max} distributions of the selected ensemble that fits the experimental SAXS data. **(a) R_g and (b) D_{max} distribution from the EOM analysis in 50 mM Glu and 50 mM Arg. (c) R_g and (d) D_{max} distributions from the EOM analysis in 6 M GdmCl. (e) R_g and (f) D_{max} distributions from the EOM analysis in 1 M TMAO. (g) Members of the conformational ensemble.** The cartoon models show some of the selected models at varying R_g values, highlighting the presence of residual helical structures. The chains are colored from the N-terminal (blue) to the C-terminal (red). The numbers refer to their position in the distribution shown in Figure 6a.

of the selected ensemble was reproducible. Also, a similar bimodal distribution was obtained independently by using another initial ensemble of 10,000 conformers devoid of any explicit secondary structure generated with the program RanCh (Bernado, Mylonas et al. 2007) (data not shown). A simulation indicated that, at least under some conditions, the program EOM is capable of revealing the existence of two different populations of different average size (Figure S2, Supplementary material). The EOM approach thus revealed features in the disordered ensemble that could not be detected from the average dimensions measured as R_S by SEC or as R_g by SAXS. In our case, the overall ensemble has average dimensions close to those calculated from random-coil statistics (Fitzkee and Rose 2004; Kohn, Millett et al. 2004; Bernado and Blackledge 2009), whereas the EOM approach suggested the co-existence of two different sub-ensembles with molecular dimensions lower and higher than those predicted for a random-coil.

The addition of stabilizing or destabilizing co-solvents induced variations in the average helical content as shown by far-UV circular dichroism spectroscopy (Figure 4). In 1 M trimethylamine-N-oxyde (TMAO), a well-known stabilizing co-solute (Bolen and Baskakov 2001), the ellipticity at 222 nm indicated a higher helical content (15 %), whereas in 6 M guanidinium chloride (GdmCl), it indicated a lower helical content (< 1 %). These co-solutes also affected the average size of P_{NTD}, and analysis of the SAXS profiles yielded a higher average R_g value of 2.9 ± 0.1 nm in 6 M GdmCl and a lower average R_g value of 2.3 ± 0.1 nm in 1 M TMAO (Table 1). The back-calculated SAXS curve from the ensemble of twenty conformers selected by the program EOM also adequately reproduced the experimental curve in the presence of GdmCl or TMAO (Supplementary material, Figure S1). The modeling of SAXS curves using the program EOM revealed that the variations of the average R_g value arose from changes in the sub-populations of compact and expanded conformers, and from slight variations of the average R_g values of these two sub-populations. In the presence of 6 M GdmCl, the sub-population of compact conformers completely disappeared, and the width of the selected distributions was significantly reduced (Figures 6c and 6d) as compared to those in 50 mM Arg/ 50 mM Glu (Figures 6a and 6b). A single maximum was observed in both distributions of R_g and D_{max} (Figure 6c and 6d) with a maximum frequency at R_g value near 2.7 nm and D_{max} value near 8.2 nm. These values were slightly lower than the R_g and D_{max} values found for the extended sub-population observed in the presence of 50 mM Arg/ 50 mM Glu, suggesting that the sub-ensemble with the highest average R_g value contained extended conformers that could not be further unfolded by addition of 6 M GdmCl. The R_g value of 2.9 nm found for this sub-ensemble is larger than that of 2.5 nm predicted for a random coil of this size in a good solvent (Fitzkee and Rose 2004) and than that of 2.3 for an IDP of this size (Bernado and Blackledge 2009). Because P_{NTD} has a high negative net charge at neutral pH, this discrepancy could result from charge repulsion leading to an expansion of the molecule (Muller-Späth, Soranno et al. 2010). GdmCl can act as both denaturant and salt, and therefore could reduce repulsions between charged residues by screening charges, a behavior which would explain the little compaction of the structural ensemble observed in 6 M GdmCl. By contrast, in the presence of 1 M TMAO, both sub-ensembles remained equally populated. They became slightly more compact than in 50 mM Arg/ 50 mM Glu with maximal frequencies centered on R_g values of 1.9 nm and 2.7 nm in agreement with the smaller experimental average R_g and D_{max} values measured directly from the SAXS curves (Table 1). The distribution of D_{max} values in the selected ensemble also appeared slightly bimodal with a maximum at a D_{max} value of 8.2 nm and a shoulder near 6.2 nm. Also, in the presence of 1 M TMAO, more models containing residual α-helical structure and models containing longer helices (the average helical content in the selected ensemble was 30 %) were selected by the EOM procedure (data not shown) in agreement with CD experiments.

DISCUSSION

The conformational ensemble representing P_{NTD}

The narrow line widths in the NMR spectrum and the limited chemical shift dispersion reflect the rapid interconversion between heterogeneous conformations of VSV P_{NTD}. ^{15}N relaxation experiments show that the protein is highly flexible with the exception of two regions, aa 2-12 and aa 25-38, that exhibit an increased level of order. This observation is supported by secondary chemical shifts that indicate that both

regions transiently populate α-helical conformations. The presence of α-helix is confirmed by circular dichroism spectrum, and the SSP parameter derived from the C_α and C_β chemical shifts provides an estimate of the time and ensemble average population of these helical elements near 9 %. Short elements of secondary structure such as α-helices, β-strands or turns are locally encoded and often stable in water (Baldwin and Rose 1999; Dyson and Wright 2002; Fitzkee, Fleming et al. 2005). α-helices form very quickly in the 10-100 nanosecond time range in agreement with the timescale probed by ^{15}N relaxation measurements (Eaton, Munoz et al. 2000). NMR experiments thus provide a dynamic picture of P_{NTD} where the polypeptide chain is highly flexible but populates fluctuating α-helices in two regions.

The hydrodynamic radius of P_{NTD} measured by SEC and its radius of gyration measured by SAXS are close to those expected for denatured proteins (Kohn, Millett et al. 2004) or for intrinsically disordered proteins (Bernado and Blackledge 2009) of the same size. The dimensions of both classes of proteins follow power-law dependences on chain length characteristic of random coil behavior (Flory 1953; Tanford 1968). However, the random coil model is insensitive to the presence of stiff elements in an otherwise flexible chain, and the presence of transient formation of α-helical elements in a disordered polypeptide chain is coherent with random-coil statistics (Millett, Doniach et al. 2002; Fitzkee and Rose 2004; Jha, Colubri et al. 2005), as already pointed out by Tanford (Tanford 1968) and demonstrated by simulations (Miller and Goebel 1968; Fitzkee and Rose 2004). The size distribution (R_g) of our selected ensemble has a width similar to that of the initial ensemble, showing that P_{NTD} samples a global conformational space predicted for a random coil although it contains on average 9% of fluctuating α-helices.

Recently, using NMR and SAXS experiments, unfolded or intrinsically disordered proteins have been modeled as structural ensembles in order to convey the highly dynamic character of these proteins (Bernado, Blanchard et al. 2005; Bernado, Mylonas et al. 2007; Jensen, Houben et al. 2008; Lange, Lakomek et al. 2008; Jensen, Salmon et al. 2010). Here, we combined information from NMR experiments showing the existence of two fluctuating α-helical elements with SAXS data to build an ensemble of 20 different conformations that together represent the dynamic structural organization of VSV P_{NTD}. The ensemble of 20 conformations selected by fitting the SAXS curve exhibits an average helical content in agreement with estimations obtained by CD and NMR and correctly captures the dynamical behavior of the polypeptide chain. The structure of the selected models must, however, be considered with care. The models selected by the EOM procedure only represent plausible conformers that together reproduce the SAXS data. They should by no means be taken as actual structures. No correlation has been found between the presence of helical conformation and the size of the conformers, and residual helical structures are found in conformers of different compactness. This argues that α-helices in P_{NTD} are stabilized by local interactions as suggested by the predicted and observed propensity of the amino acid sequence to form helices in free solution and, therefore, their presence is independent of the compactness of the peptide chain. However, it is probable that the information content of SAXS data is insufficient to correlate the presence of helical structure with the degree of compaction of the chain.

In addition, a detailed analysis of the population distribution in the EOM-selected ensemble and the effects of denaturing and stabilizing co-solutes suggest the co-existence of two sub-populations with different degrees of compactness. A simulation indicates that EOM is capable of distinguishing between unimodal and bimodal distributions under certain conditions (Figure S2, Supporting information). The addition of denaturing or stabilizing co-solutes changes the average size of the molecule and the population distribution selected by the EOM program. In particular, in 6 M GdmCl, the width of the size distribution is substantially reduced and the distribution becomes unimodal (Figure 6c), whereas in 1 M TMAO, the bimodal distribution becomes more clearly visible (Figure 6e). These results suggest the presence of one sub-population that is more compact than that predicted by random coil statistics and one that is more expanded. Some IDPs are more compact than predicted from random coil statistics because of the presence of long range interactions involving hydrophobic clusters (Felitsky, Lietzow et al. 2008; Mittag, Orlicky et al. 2008), hydrogen bonding (Moglich, Joder et al. 2006) or charge-charge attractions (Mao, Crick et al. 2010; Muller-Späth, Soranno et al. 2010). Others appear more expanded because of charge-charge repulsions (Uversky, Gillespie et al. 1999; Uversky,

Li et al. 2002; Permyakov, Millett et al. 2003). Charge-charge interactions thus seem to play a critical role in modulating the dimensions of the conformational ensemble of IDPs (Mao, Crick et al. 2010; Muller-Späth, Soranno et al. 2010); repulsions between like charges can lead to the expansion of the peptide chain, whereas attractions between opposite charges can lead to collapse. VSV P_{NTD} is rich in charged residues, containing 15 acidic residues (D and E), 7 basic residues (K and R) and 1 histidine residue (net charge at neutral pH of -7 or -8 if tehc charge on H is accounted for or not) and therefore different interactions between charged residues could explain the existence of the two different conformational sub-ensembles. The salt properties of GdmCl can explain the disappearance in 6 M GdmCl of the compact sub-ensemble if its formation is driven by opposite charge interactions. Conversely, TMAO favors compaction of random coils (Qu, Bolen et al. 1998), and we observed slight reductions of the average dimensions of both selected conformational sub-ensembles. Also, TMAO induces helix formation (Celinski and Scholtz 2002), and we observed a larger α-helical content as judged by CD spectroscopy and a higher number of models containing helical elements in the EOM-selected ensemble. In conclusion, our data show that P_{NTD} is highly flexible and has overall molecular dimensions of a random coil chain. However, NMR and CD spectroscopies revealed the presence of fluctuating helical structures and the modeling of SAXS data suggests the co-existence of sub-populations of compact and extended conformers, which together suggest a pre-organization of isolated P_{NTD}.

Role for a MoRE in the chaperone function of P

The discovery here that, in solution, P_{NTD} populates compact and pre-organized conformers containing helical elements suggests a mechanism by which P_{NTD} could fold upon binding to its N^0 partner (Wright and Dyson 2009). P_{NTD} could thus be a molecular recognition element or MoRE, *i.e.* a short motif within a disordered protein that promotes binding to a partner (Bourhis, Johansson et al. 2004; Fuxreiter, Simon et al. 2004; Oldfield, Cheng et al. 2005; Mohan, Oldfield et al. 2006; Jensen, Houben et al. 2008). Structural ensembles of disordered proteins that fold upon binding to a partner usually contain conformations similar to those adopted in the final complex (Gsponer, Christodoulou et al. 2008; Jensen, Houben et al. 2008; Lange, Lakomek et al. 2008). The fluctuating α-helices might prefigure structural elements involved in the formation of the N^0-P complex. The first helix (aa 8-12) is amphipathic exposing three positively charged residues and two negatively charged residues on one face of the helix and four hydrophobic residues on the other face. The second helix (aa 25-38) is hydrophilic exposing charged residues around the helix and containing only two hydrophobic residues (I27 and Y37). Both helices could thus interact with positively charged residues lining the RNA binding groove of N^0. One might speculate that the highly flexible character of P_{NTD} helps the protein to get into the RNA binding groove and allows the correct positioning of the helical elements.

In the viral replication cycle, the N^0-P complex is not an end-product. It forms only transiently, as an intermediate during the synthesis of new nucleocapsids. The N^0-bound P must be outcompeted by the newly synthesized genomic RNA, and therefore the affinity of P for N^0 cannot be too high. The mechanism of folding upon binding provides a means of specific recognition without the corollary of high affinity (Dyson and Wright 2002; Wright and Dyson 2009). Indeed, the folding or the adoption of a rigid structure by the ligand when it binds to its receptor will generally lead to the formation of multiple specific intermolecular interactions that confer a great specificity, and consequently a high affinity. If the ligand is disordered in its unbound form, this strong binding energy is opposed by the high entropy of the disordered protein. We might thus speculate that by being disordered in its free form, P_{NTD} recognizes N^0 with high specificity but moderate affinity, thus allowing its displacement by the newly synthesized viral RNA.

MATERIALS AND METHODS

Cloning, Expression and Purification of VSV P_{NTD}

The cDNA encoding VSV PNTD (aa 1-60) was amplified and cloned into a pET28a(+) plasmid between the NcoI and XhoI restriction sites using standard PCR techniques. A His6-tag and including a linker of two amino-acids (Glu-Leu) was introduced at the C-terminal extremity. The construct was verified by standard dideoxy sequencing. The plasmid was transfected into *E.coli* strain BL21(DE3). ^{15}N and ^{13}C protein samples for NMR spectroscopy were produced in M9 minimal medium supplemented with 1.0 $g.L^{-1}$ of $^{15}NH_4Cl$, 2.0 $g.L^{-1}$ of ^{13}C glucose, and MEM vitamins (Gibco). Cells were grown in the medium containing 100 µg/mL kanamycin at 37°C until

O.D600 reached 0.6-0.8 A.U. At this point, IPTG at a final concentration of 0.5 mM was added. After an incubation of 3 hours, cells were harvested by centrifugation and resuspended in 20 mM Tris-HCl, 150 mM NaCl, 50 mM Glu, 50 mM Arg, pH 7.5 (Buffer A) supplemented with anti-proteases (Complete™ Protease Inhibitor Cocktail Tablets, Roche) and cells were disrupted by sonication. The extract was centrifuged at 20,000 g during 30 min at 4°C. The supernatant was filtered (0.45 µm) and loaded onto a Ni^{2+} resin column, pre-equilibrated with Buffer A. The resin was washed with three bed volumes of Buffer A, then with three bed volumes of 20 mM Tris-HCl, 1.5 M NaCl, 10 mM imidazole at pH 7.5, and the protein was eluted using Buffer A supplemented with 400 mM imidazole. The recombinant protein was loaded onto a size exclusion chromatography column (HiLoad 16/60 Superdex 75 prep grade, GE Healthcare) pre-equilibrated with Buffer A. Separations were performed at a flow rate of 0.5 mL/min. The purified protein was concentrated (Vivaspin - 5,000 MWCO) and stored at 4°C. The purity of the protein samples was assessed by SDS-PAGE. Protein concentration was measured by refractometry.

NMR spectroscopy

The resonance assignment of VSV P_{NTD} was carried out using a double-labeled (^{13}C, ^{15}N) sample of the peptide in 20 mM Tris-HCl, 150 mM NaCl, 50 mM Glu, 50 mM Arg with 10% D_2O adjusted to pH 6.0 to avoid loss of resonances due to fast amide proton exchange with the solvent protons. The protein concentration was 0.9 mM. All NMR experiments were carried out at 10°C and a ^1H resonance frequency of 599.67 MHz. The assignment was obtained from a series of BEST-type triple resonance experiments (Schanda, Van Melckebeke et al. 2006; Lescop, Schanda et al. 2007): HNCO, intra-residue HN(CA)CO, HN(CO)CA, intra-residue HNCA, HN(COCA)CB and intra-residue HN(CA)CB. The spectra were acquired with a sweep width of 8 kHz and 512 complex points in the ^1H dimension, and a sweep width of 1.2 kHz and 32 complex points in the ^{15}N dimension. For the ^{13}C dimension, the spectra were acquired with a sweep width of 1.2 kHz and 60 complex points (HNCO and intra-residue HN(CA)CO), 3 kHz and 110 complex points (HN(CO)CA and intra-residue HNCA), and 10 kHz and 100 complex points (HN(COCA)CB and intra-residue HN(CA)CB). All spectra were processed with NMRPipe (Delaglio, Grzesiek et al. 1995) and analyzed using SPARKY (Goddard and Kneller 2003). Automatic assignment of the resonances on the basis of the SPARKY peak lists was done using the program MARS (Jung and Zweckstetter 2004).

^{15}N R_1, R_2 (CPMG) relaxation and {^1H}-^{15}N steady-state nuclear Overhauser effect (nOe) experiments were acquired using standard pulse sequences (Farrow, Muhandiram et al. 1994) on an ^{15}N labeled sample of P_{NTD} with a protein concentration of 1.4 mM. The buffer conditions were as described above for the double-labeled sample. The spectra were acquired with a sweep width of 7.0 kHz and 512 complex points in the ^1H dimension, and a sweep width of 1.2 kHz and 250 complex points in the ^{15}N dimension.

The magnetization decay was sampled at (0, 100, 200, 400, 600, 800, 1100, 1500 and 1900) ms for longitudinal and at (10, 30, 50, 70, 90, 130, 170, 210 and 250) ms for transverse relaxation and the peak heights were used to extract the relaxation rates. To obtain estimates of the errors on the relaxation rates, a repeat measurement of one of the relaxation delays was performed in each case (600 ms for R_1 and 70 ms for R_2).

For the heteronuclear nOes, the amide protons were saturated using a 3 s WALTZ16 decoupling scheme that in the reference experiment was replaced by a 3 s delay. The recycle delay in both experiments was set to 2 s. Heteronuclear NOE values were obtained from the ratio between signal intensities in the saturated and the reference experiments, where the standard deviation in the noise was taken as a measure for the error in the signal intensity. The experiment was repeated twice and the NOE values were averaged.

SEC-MALLS-RI

Size-exclusion chromatography (SEC) was performed with an S75 Superdex column (GE Healthcare) equilibrated with 20 mM Tris/HCl at pH 7.5 containing 150 mM NaCl. Separations were performed at 20°C with a flow rate of 0.5 $ml.min^{-1}$. 50 µL of a protein solution at a concentration of 470 µM (3.8 $mg.mL^{-1}$) was injected. On-line Multi-Angle Laser Light Scattering (MALLS) detection was performed with a DAWN-EOS detector (Wyatt Technology Corp., Santa Barbara, CA) using a laser emitting at 690 nm. Data were analyzed and weight-averaged molecular masses (M_w) were calculated using the ASTRA software (Wyatt Technology Corp., Santa Barbara, CA) as described previously (Gérard, Ribeiro et al. 2007). The column was calibrated with proteins of known Stokes' radii (R_S) (35): bovine serum albumine (R_S = 3.4 nm), RNAse A (R_S = 1.9 nm), ovalbumin (R_S = 3.0 nm), β-lactoglobulin (R_S = 2.7 nm), and chymotrypsinogen (R_S = 2.3 nm) (Uversky 1993).

Circular dichroism spectroscopy

Far-UV circular dichroism spectra were recorded at 20°C on a JASCO model J-810 CD spectropolarimeter equipped with a Peltier temperature controller. VSV P_{NTD} was diluted to final a concentration of about 950 µM, in 20 mM Tris-HCl pH 7.5 containing 150 mM NaCl, 50 mM Arg, 50 mM Glu. Spectra were measured in a cuvette with a path length of 0.1 mm. After subtracting the blank signal, the CD signal (in millidegrees) was converted to mean molar residue ellipticity (in $deg.cm^2.dmol^{-1}$) and helix content was calculated by using formalism derived from isolated peptide (Luo and Baldwin 1997) as described previously (Gérard, Ribeiro et al. 2009).

Small-angle X-ray scattering (SAXS) experiments

The monodispersity of the samples used in SAXS experiments was checked by size exclusion chromatography (SEC) combined with detection by multi-angle laser light scattering (MALLS) and refractometry (RI) (Wyatt 1998; Gérard, Ribeiro et al. 2007). SAXS data were collected at the European Synchrotron Radiation Facility (Grenoble, France) on beamline ID14-3. The sample-to-detector distance was 1 m and the wavelength of the X-rays was 0.0995 nm. Samples were contained in a 1.9 mm wide quartz capillary. The time of exposure was optimized for reducing radiation damage. Data acquisition was performed at 20 °C. Protein concentrations ranged from 5 to 18 $mg.ml^{-1}$. Data reduction was performed using the established procedure available at ID14-3, and buffer background runs were subtracted from sample runs.

The radius of gyration and forward intensity at zero angle ($I(0)$) were determined with the programs PRIMUS (Konarev, Volkov et al. 2003) by using the Guinier approximation at low scattering vector values, Q ($Q = \frac{4\pi \sin\theta}{\lambda}$), in a $Q R_g$ range up to 1.3:

$$\ln I(Q) = \ln I(0) - \frac{R_g^2 Q^2}{3} \quad \text{(Eq. 1)}$$

The forward scattering intensity was calibrated using bovine serum albumin and lysozyme as references. The radius of gyration and pairwise distance distribution function, *P(r)*, were calculated with the program GNOM (Semenyuk and Svergun 1991). The maximum dimension (D_{max}) value was adjusted so that the R_g value obtained from GNOM agreed with that obtained from the Guinier analysis.

The radius of gyration, R_g, was calculated using the power-law dependence established for random-coil polymers (Flory 1953):

$$R_g = R_0 N^{\nu} \quad \text{(Eq. 2)}$$

where R_0 is the persistence length, N is the number of residues in the peptide and ν is the exponential scaling factor. Values for R_0 and ν were taken either from the random-coil polymer theory (Fitzkee and Rose 2004) or from experiments with IDPs (Bernado and Blackledge 2009).

A conformational ensemble of 4,500 conformers was generated with the program Flexible-Meccano (Bernado, Blanchard et al. 2005) by pooling nine sub-ensembles of 500 conformers in which helical element of different lengths were imposed in the putative helical regions of P_{NTD} identified by NMR spectroscopy as follows: Ensemble 1: no helix; Ensemble 2: aa 2-9; Ensemble 3: aa 2-19; Ensemble 4: aa 25-32; Ensemble 5: aa 25-32; Ensemble 6: aa 2-9 and aa 25-32; Ensemble 7: aa 2-9 and aa 25-38; Ensemble 8: aa 2-19 and aa 25-32; Ensemble 9: aa 2-19 and aa 25-38. All helices were invoked with a population of 100%. Side chains for these conformations were added with the program SCCOMP (Eyal, Najmanovich et al. 2004). An optimized ensemble of conformations that agrees with the experimental SAXS data was selected from the large conformational ensemble using the Ensemble Optimization Method (EOM) software (Bernado, Mylonas et al. 2007). The size of the optimized ensemble was varied from 50 to 1. Average chemical shifts for the selected ensemble were calculated with the program SPARTA (Shen and Bax 2007).

Acknowledgments

This work was supported by grants from the French ANR (ANR-07-001-01 (ANRAGE)), the FINOVI foundation and Lyonbiopôle. E.A.R. was supported by postdoctoral fellowships from both the ANR and the FINOVI programs. C.L. and F.G. were supported by MENRT fellowships from the French government. M.R.J. received a long-term EMBO fellowship and was supported by Lundbeckfonden, Denmark. We thank the Partnership for Structural Biology for the excellent structural biology environment.

REFERENCES

1. Wright, P.E. and H.J. Dyson, *Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm.* J Mol Biol, 1999. **293**(2): p. 321-31.
2. Dunker, A.K., C.J. Brown, and Z. Obradovic, *Identification and functions of usefully disordered proteins.* Adv Protein Chem, 2002. **62**: p. 25-49.
3. Wright, P.E. and H.J. Dyson, *Linking folding and binding.* Curr Opin Struct Biol, 2009. **19**(1): p. 31-8.
4. Tompa, P., *Intrinsically unstructured proteins.* Trends Biochem Sci, 2002. **27**(10): p. 527-33.
5. Uversky, V.N., J.R. Gillespie, and A.L. Fink, *Why are "natively unfolded" proteins unstructured under physiologic conditions?* Proteins, 2000. **41**(3): p. 415-27.
6. Eliezer, D., *Biophysical characterization of intrinsically disordered proteins.* Curr Opin Struct Biol, 2009. **19**(1): p. 23-30.
7. Dyson, H.J. and P.E. Wright, *Unfolded proteins and protein folding studied by NMR.* Chem Rev, 2004. **104**(8): p. 3607-22.
8. Jensen, M.R., et al., *Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings.* Structure, 2009. **17**(9): p. 1169-85.
9. Bernado, P., et al., *Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering.* J Am Chem Soc, 2007.
10. Bernado, P. and M. Blackledge, *A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering.* Biophys J, 2009. **97**(10): p. 2839-45.
11. He, B., et al., *Predicting intrinsic disorder in proteins: an overview.* Cell Res, 2009. **19**(8): p. 929-49.
12. Gérard, F.C.A., et al., *Modular organization of rabies virus phosphoprotein.* J. Mol. Biol., 2009. **388**: p. 978-996.
13. Ward, J.J., et al., *Prediction and functional analysis of native disorder in proteins from the three kingdoms of life.* J Mol Biol, 2004. **337**(3): p. 635-45.
14. Oldfield, C.J., et al., *Comparing and combining predictors of mostly disordered proteins.* Biochemistry, 2005. **44**(6): p. 1989-2000.
15. Dunker, A.K., et al., *Intrinsic protein disorder in complete genomes.* Genome Inform Ser Workshop Genome Inform, 2000. **11**: p. 161-71.
16. Tokuriki, N., et al., *Do viral proteins possess unique biophysical features?* Trends Biochem Sci, 2009. **34**(2): p. 53-9.
17. Fields, B.N., et al., *Fields' virology, 3rd edn.* 1996, New York: Lippincott-Raven Publishers.
18. Arnheiter, H., et al., *Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis.* Cell, 1985. **41**(1): p. 259-67.
19. Albertini, A.A., et al., *Crystal structure of the rabies virus nucleoprotein-RNA complex.* Science, 2006. **313**(5785): p. 360-3.
20. Green, T.J., et al., *Structure of the vesicular stomatitis virus nucleoprotein-RNA complex.* Science, 2006. **313**(5785): p. 357-60.
21. Emerson, S.U. and Y. Yu, *Both NS and L proteins are required for in vitro RNA synthesis by vesicular stomatitis virus.* J Virol, 1975. **15**(6): p. 1348-56.
22. Masters, P.S. and A.K. Banerjee, *Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA.* J Virol, 1988. **62**(8): p. 2658-64.
23. Emerson, S.U. and M. Schubert, *Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus.* Proc Natl Acad Sci U S A, 1987. **84**(16): p. 5655-9.
24. Mellon, M.G. and S.U. Emerson, *Rebinding of transcriptase components (L and NS proteins) to the nucleocapsid template of vesicular stomatitis virus.* J Virol, 1978. **27**(3): p. 560-7.
25. Peluso, R.W. and S.A. Moyer, *Viral proteins required for the in vitro replication of vesicular stomatitis virus defective interfering particle genome RNA.* Virology, 1988. **162**(2): p. 369-76.
26. Peluso, R.W., *Kinetic, quantitative, and functional analysis of multiple forms of the vesicular stomatitis virus nucleocapsid protein in infected cells.* J Virol, 1988. **62**(8): p. 2799-807.
27. Howard, M. and G. Wertz, *Vesicular stomatitis virus RNA replication: a role for the NS protein.* J Gen Virol, 1989. **70 (Pt 10)**: p. 2683-94.

28. Mavrakis, M., et al., *Isolation and characterisation of the rabies virus N degrees-P complex produced in insect cells.* Virology, 2003. **305**(2): p. 406-14.
29. Mavrakis, M., et al., *Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA.* Virology, 2006. **349**(2): p. 422-9.
30. Chen, M., T. Ogino, and A.K. Banerjee, *Interaction of vesicular stomatitis virus P and N proteins: Identification of two overlapping domains at the N-terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA.* J Virol, 2007. **81**: p. 13478-13485.
31. Hudson, L.D., C. Condra, and R.A. Lazzarini, *Cloning and expression of a viral phosphoprotein: structure suggests vesicular stomatitis virus NS may function by mimicking an RNA template.* J Gen Virol, 1986. **67 (Pt 8)**: p. 1571-9.
32. Gérard, F., et al., *Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution.* Biochemistry, 2007. **46**: p. 10328-10338.
33. Ribeiro, E.A., Jr., et al., *Solution Structure of the C-Terminal Nucleoprotein-RNA Binding Domain of the Vesicular Stomatitis Virus Phosphoprotein.* J Mol Biol, 2008. **382**: p. 525-538.
34. Dyson, H.J. and P.E. Wright, *Intrinsically unstructured proteins and their functions.* Nat Rev Mol Cell Biol, 2005. **6**(3): p. 197-208.
35. Sugase, K., H.J. Dyson, and P.E. Wright, *Mechanism of coupled folding and binding of an intrinsically disordered protein.* Nature, 2007. **447**: p. 920-921.
36. Jensen, M.R., et al., *Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein.* J Am Chem Soc, 2008. **130**(25): p. 8055-61.
37. Mohan, A., et al., *Analysis of molecular recognition features (MoRFs).* J Mol Biol, 2006. **362**(5): p. 1043-59.
38. Golovanov, A.P., et al., *A simple method for improving protein solubility and long-term stability.* J Am Chem Soc, 2004. **126**(29): p. 8933-9.
39. Uversky, V.N., *Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule.* Biochemistry, 1993. **32**(48): p. 13288-98.
40. Fitzkee, N.C. and G.D. Rose, *Reassessing random-coil statistics in unfolded proteins.* Proc Natl Acad Sci U S A, 2004. **101**(34): p. 12497-502.
41. Luo, P. and R.L. Baldwin, *Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water.* Biochemistry, 1997. **36**: p. 8413-8421.
42. Wishart, D.S., B.D. Sykes, and F.M. Richards, *Relationship between nuclear magnetic resonance chemical shift and protein secondary structure.* J Mol Biol, 1991. **222**(2): p. 311-33.
43. Jensen, M.R., et al., *Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts.* J Am Chem Soc, 2010. **132**(4): p. 1270-2.
44. Marsh, J.A., et al., *Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation.* Protein Sci, 2006. **15**(12): p. 2795-804.
45. Eliezer, D., et al., *Native and non-native secondary structure and dynamics in the pH 4 intermediate of apomyoglobin.* Biochemistry, 2000. **39**(11): p. 2894-901.
46. Spera, S. and A. Bax, *Empirical Correlation between Protein Backbone Conformation and Ca and Cb ^{13}C Nuclear Magnetic Resonance Chemical Shifts.* J Am Chem Soc, 1991. **113**: p. 5490-5492.
47. Munoz, V. and L. Serrano, *Elucidating the folding problem of helical peptides using empirical parameters.* Nat Struct Biol, 1994. **1**(6): p. 399-409.
48. Lacroix, E., A.R. Viguera, and L. Serrano, *Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters.* J Mol Biol, 1998. **284**(1): p. 173-91.
49. Doig, A.J. and R.L. Baldwin, *N- and C-capping preferences for all 20 amino acids in alpha-helical peptides.* Protein Sci, 1995. **4**(7): p. 1325-36.
50. Franke, D. and D.I. Svergun, *DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering.* J. Appl. Cryst., 2009. **42**: p. 342-346.
51. Svergun, D.I., M.V. Petoukhov, and M.H. Koch, *Determination of domain structure of proteins from X-ray solution scattering.* Biophys J, 2001. **80**(6): p. 2946-53.
52. Konarev, P., et al., *ATSAS 2.1, a program package for small-angle scattering data analysis.* J. Appl. Cryst., 2006. **39**: p. 277-286.
53. Putnam, C.D., et al., *X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution.* Q Rev Biophys, 2007. **40**(3): p. 191-285.
54. Bernado, P., et al., *A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering.* Proc Natl Acad Sci U S A, 2005. **102**(47): p. 17002-7.
55. Kohn, J.E., et al., *Random-coil behavior and the dimensions of chemically unfolded proteins.* Proc Natl Acad Sci U S A, 2004. **101**(34): p. 12491-6.
56. Bolen, D.W. and I.V. Baskakov, *The osmophobic effect: natural selection of a thermodynamic force in protein folding.* J Mol Biol, 2001. **310**(5): p. 955-63.
57. Muller-Späth, S., et al., *Charge interactions can dominate the dimensions of intrinsically disordered proteins.* Proc Natl Acad Sci U S A, 2010. **107**: p. 14609-14614.
58. Baldwin, R.L. and G.D. Rose, *Is protein folding hierarchic? I. Local structure and peptide folding.* Trends Biochem Sci, 1999. **24**(1): p. 26-33.
59. Fitzkee, N.C., et al., *Are proteins made from a limited parts list?* Trends Biochem Sci, 2005. **30**(2): p. 73-80.
60. Dyson, H.J. and P.E. Wright, *Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance.* Adv Protein Chem, 2002. **62**: p. 311-40.
61. Eaton, W.A., et al., *Fast kinetics and mechanisms in protein folding.* Annu Rev Biophys Biomol Struct, 2000. **29**: p. 327-59.
62. Flory, P.J., *Principles of polymer chemistry*. 1953, Ithaca, N.Y.: Cornell University Press.
63. Tanford, C., *Protein denaturation.* Adv Protein Chem, 1968. **23**: p. 121-282.
64. Millett, I.S., S. Doniach, and K.W. Plaxco, *Toward a taxonomy of the denatured state: small angle scattering studies of unfolded proteins.* Adv Protein Chem, 2002. **62**: p. 241-62.
65. Jha, A.K., et al., *Statistical coil model of the unfolded state: resolving the reconciliation problem.* Proc Natl Acad Sci U S A, 2005. **102**(37): p. 13099-104.
66. Miller, W.G. and C.V. Goebel, *Dimensions of protein random coils.* Biochemistry, 1968. **7**(11): p. 3925-35.
67. Lange, O.F., et al., *Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution.* Science, 2008. **320**(5882): p. 1471-5.
68. Felitsky, D.J., et al., *Modeling transient collapsed states of an unfolded protein to provide insights into early folding events.* Proc Natl Acad Sci U S A, 2008. **105**(17): p. 6278-83.

69. Mittag, T., et al., *Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor.* Proc Natl Acad Sci U S A, 2008. **105**(46): p. 17772-7.
70. Moglich, A., K. Joder, and T. Kiefhaber, *End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation.* Proc Natl Acad Sci U S A, 2006. **103**(33): p. 12394-9.
71. Mao, A.H., et al., *Net charge per residue modulates conformational ensembles of intrinsically disordered proteins.* Proc Natl Acad Sci U S A, 2010. **107**(18): p. 8183-8.
72. Uversky, V.N., et al., *Natively unfolded human prothymosin alpha adopts partially folded collapsed conformation at acidic pH.* Biochemistry, 1999. **38**(45): p. 15009-16.
73. Permyakov, S.E., et al., *Natively unfolded C-terminal domain of caldesmon remains substantially unstructured after the effective binding to calmodulin.* Proteins, 2003. **53**(4): p. 855-62.
74. Uversky, V.N., et al., *Biophysical properties of the synucleins and their propensities to fibrillate: inhibition of alpha-synuclein assembly by beta- and gamma-synucleins.* J Biol Chem, 2002. **277**(14): p. 11970-8.
75. Qu, Y., C.L. Bolen, and D.W. Bolen, *Osmolyte-driven contraction of a random coil protein.* Proc Natl Acad Sci U S A, 1998. **95**(16): p. 9268-73.
76. Celinski, S.A. and J.M. Scholtz, *Osmolyte effects on helix formation in peptides and the stability of coiled-coils.* Protein Sci, 2002. **11**(8): p. 2048-51.
77. Bourhis, J.M., et al., *The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner.* Virus Res, 2004. **99**(2): p. 157-67.
78. Oldfield, C.J., et al., *Coupled folding and binding with alpha-helix-forming molecular recognition elements.* Biochemistry, 2005. **44**(37): p. 12454-70.
79. Fuxreiter, M., et al., *Preformed structural elements feature in partner recognition by intrinsically unstructured proteins.* J Mol Biol, 2004. **338**(5): p. 1015-26.
80. Gsponer, J., et al., *A coupled equilibrium shift mechanism in calmodulin-mediated signal transduction.* Structure, 2008. **16**(5): p. 736-46.
81. Dyson, H.J. and P.E. Wright, *Coupling of folding and binding for unstructured proteins.* Curr Opin Struct Biol, 2002. **12**(1): p. 54-60.
82. Schanda, P., H. Van Melckebeke, and B. Brutscher, *Speeding up three-dimensional protein NMR experiments to a few minutes.* J Am Chem Soc, 2006. **128**(28): p. 9042-3.
83. Lescop, E., P. Schanda, and B. Brutscher, *A set of BEST triple-resonance experiments for time-optimized protein resonance assignment.* J Magn Reson, 2007. **187**(1): p. 163-9.
84. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J Biomol NMR, 1995. **6**(3): p. 277-93.
85. Goddard, T.D. and D.G. Kneller, *SPARKY 3*. 2003: University of California, San Francisco.
86. Jung, Y.S. and M. Zweckstetter, *Mars – robust automatic backbone assignment of proteins.* J. Biomolecular NMR, 2004. **30**: p. 11-23.
87. Farrow, N.A., et al., *Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation.* Biochemistry, 1994. **33**(19): p. 5984-6003.
88. Wyatt, P.J., *Submicrometer Particle Sizing by Multiangle Light Scattering following Fractionation.* J Colloid Interface Sci, 1998. **197**(1): p. 9-20.
89. Konarev, P.V., et al., *PRIMUS: a Windows PC-based system for small-angle scattering data analysis.* J. Appl. Cryst., 2003. **36**: p. 1277-1282.
90. Semenyuk, A.V. and D. Svergun, *GNOM - a program package for small-angle scattering data processing.* J. Appl. Crystallog., 1991. **24**: p. 537-540.
91. Eyal, E., et al., *Importance of solvent accessibility and contact surfaces in modeling side-chain conformations in proteins.* J Comput Chem, 2004. **25**(5): p. 712-24.
92. Shen, Y. and A. Bax, *Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology.* J Biomol NMR, 2007. **38**(4): p. 289-302.

ELECTRONIC SUPPLEMENTARY MATERIAL

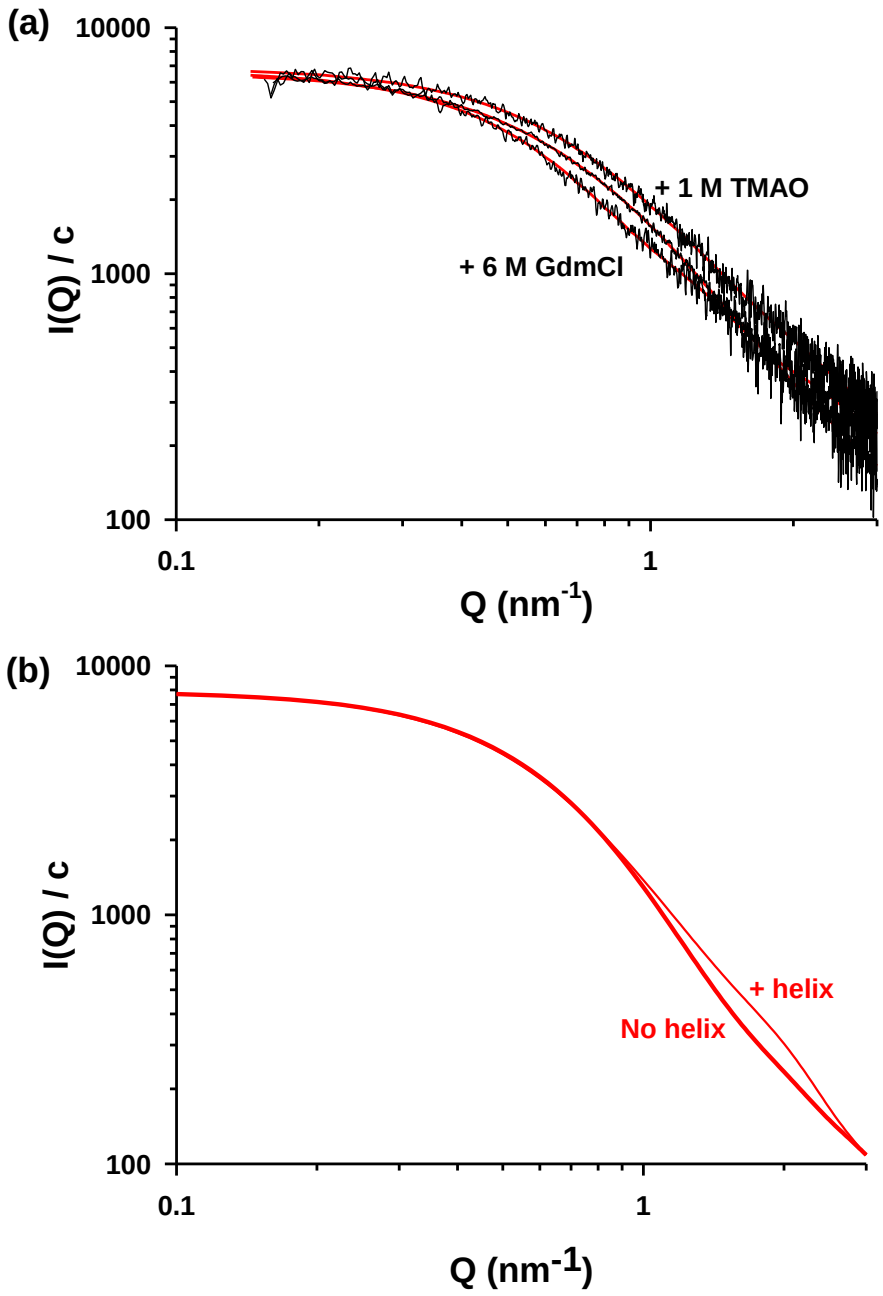


Figure S1. (a) Fitting of SAXS curve recorded for P_{NTD}. The SAXS curves obtained in 50 mM Arg, 50 mM Glu at three different protein concentrations (Figure 2) were merged and the EOM analysis was performed on the merged curve. In 1 M TMAO and 6 M GdmCl, SAXS curve were obtained at 1 mg.mL^{-1}. The red curve shows the fitted curve obtained using the EOM procedure. Data were recorded at 20°C in 20 mM Tri-HCl pH 7.5 containing 50mM Arg and 50 mM Glu, 1 M TMAO or 6 M GdmCl. **(b) SAXS curves of ensemble containing different amounts of helical elements.** The curves are average SAXS profiles calculated for the 500 models that contains no helical elements (no helix) and for the 500 models that contains helices in the regions 2-19 and 25-38 (32 % helical content).

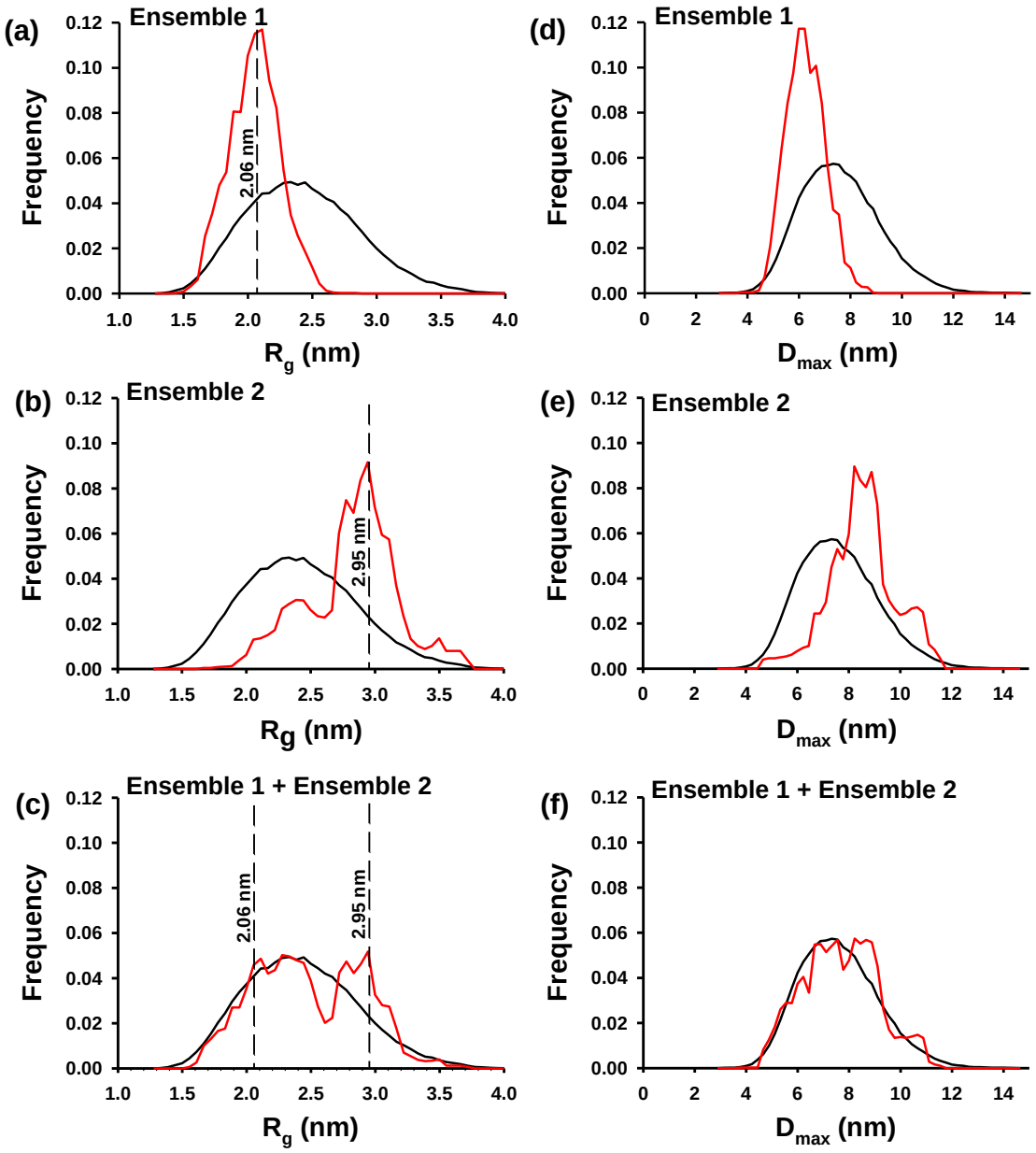


Figure S2. Simulation of an EOM analysis with mocked P_{NTD} ensembles. To assess the capacity of the EOM procedure for revealing the existence of two different populations of different average size, we performed a simulation. We randomly generated two ensembles containing three equally-populated compact conformers for ensemble 1 and three equally-populated extended conformers for ensemble 2, and calculated a SAXS profile for each ensemble. We also mixed both ensembles and calculated a SAXS profile for the overall ensemble. We then used the program EOM for analyzing these theoretical SAXS curves using our initial ensemble of conformers. Analysis of the curves generated with a single mocked ensemble, yielded EOM-selected ensembles that had a narrow distribution of R_g and D_{max} values centered on the average R_g (Panels S2a and S2b) and D_{max} values (Panels S2d and S2e) of the theoretical ensemble. **(a, d) Unimodal theoretical ensemble composed of three compact structures.** The black curve shows the R_g and D_{max} distributions calculated for the initial ensemble, whereas the red curve shows the R_g and D_{max} distributions for the selected ensemble that collectively reproduces the theoretical SAXS curve. The dashed line indicates the average Rg value of the theoretical ensemble (R_g values for ensemble 1: 1.94, 2.07 and 2.16 nm. Average Rg value: 2.06 nm). **(b, e) Unimodal theoretical ensemble composed of three extended structures.** Same legend as in panel a and d. R_g values for ensemble 2: 2.86, 2.89 and 3.09 nm. Average value: 2.95 nm. **(c) Bimodal theoretical ensemble composed of the two ensembles.** Analysis of the curve generated with the pooled ensembles yielded an EOM-selected ensemble with a broader distribution of R_g values that appears bimodal (Panel 6c). The distribution of D_{max} values is also broader but is not bimodal (Panel 6f). These calculations suggest that the information content of the SAXS curves is sufficient, at least in some conditions, for allowing the EOM program to distinguish between unimodal and bimodal distributions and for showing the co-existence of compact and extended populations of conformers.

III. ARTICLE III: α-Helical and Polyproline II Structure in The Intrinsically Disordered N^0 Binding Region of Rabies Virus Phosphoprotein: Insights from Small Angle X-ray Scattering

Auteurs: Cédric Leyrat, Euripedes de Almeida Ribeiro Jr, Louiza Zerrad, Malene Ringkjøbing Jensen, Rob Ruigrok, Martin Blackledge et Marc Jamin.

En préparation

α-Helical and Polyproline II Structure in The Intrinsically Disordered N^0 Binding Region of Rabies Virus Phosphoprotein: Insights from Small Angle X-ray Scattering

Cédric Leyrat[1], Euripedes de Almeida Ribeiro Jr[1], Louiza Zerrad[3], Malene Ringkjøbing Jensen[2], Rob Ruigrok[1], Martin Blackledge[2] and Marc Jamin[1*]

[1]Unit of Virus Host Cell Interactions, UMI 3265 UJF-EMBL-CNRS
6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France
[2]Protein Dynamics and Flexibility, Institut de Biologie Structurale, UMR 5075 CEA-CNRS-UJF, 41 rue Jules Horowitz, Grenoble 38027, France
[3]Structural Biology Group, European Synchrotron Radiation Facility, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

Keywords: phosphoprotein, intrinsically disordered regions

* **Corresponding author:** Unit of Virus Host Cell Interactions (UVHCI), UMI 3265 UJF-EMBL-CNRS, 6, rue Jules Horowitz, B.P. 181, 38042 Grenoble Cedex 9, France

E-mail : jamin@embl.fr, Phone: + 33 4 76 20 94 62, Fax : + 33 4 76 20 94 00

Abstract

The N-terminal region of Rabies Virus phosphoprotein comprising residues 1-68 (P_{68}) is an intrinsically disordered polypeptide that binds to the RNA-free nucleoprotein N°. Here, we show using small sngle X-ray scattering (SAXS) and the Ensemble Optimization Method (EOM) that P_{68} exists as an ensemble of conformers of different compactness. We studied the effect of protein concentration, pH, temperature, and of the presence of stabilizing or denaturing co-solutes on the ensembles populated by P_{68} in solution. We found that: (1) at high protein concentration, repulsive intra-molecular interactions favor compact states over extended ones leading to a decrease in the measured R_g and an increased frequency of α-helical conformers; (2) P_{68} aggregation at low pH and high protein concentration is correlated with an increase in the secondary structure content as observed by CD; (3) In the presence of stabilizing co-solute, such as TriMethylaAmine N-Oxide (TMAO), P_{68} populates compact conformations, whereas in the presence of guanidinium hydrochloride (Gdn-HCl) it populates extended conformations; (4) finally, temperature increase causes a gradual decrease in the Rg of the molecule measured by SAXS and affects the proportions of unfolded and partially folded conformers in the ensemble, leading both to an increase in P_{68} α-helical content and to a decrease in polyproline II structure. These results demonstrate that P_{68} possesses some degree of structural pre-organization in solution and suggest that P_{68} acts as a molecular recognition element (MoRE) that folds upon binding to its partner N^0.

Introduction

Rabies virus (RV) is an enveloped RNA virus of the genus *Lyssavirus* in the family *Rhabdoviridae*. Its non-segmented, single-stranded, negative sense RNA genome is packaged by the viral nucleoprotein (N) within a helical nucleocapsid. The N-RNA complex, rather than the naked RNA, serves as a template for both transcription and replication. These reactions are carried out by a two-subunit RNA-dependent RNA polymerase made of the large protein (L) associated with the phosphoprotein (P) (Arnheiter, Davis et al. 1985; Fields, Knipe et al. 1996). During genome replication, synthesis of viral RNA is concomitant with its encapsidation by N and therefore the continuous production of N is required. N has strong affinity for non-specific RNA and it self-assembles on cellular RNA to form nucleocapsid-like particles (Iseni, Barge et al. 1998; Schoehn, Iseni et al. 2001). A control mechanism prevents illegitimate encapsidation of cellular RNA, where P plays a major role. P is a modular protein that contains both structured domains and intrinsically

disordered regions (IDRs) (Gérard, Ribeiro et al. 2007; Gérard, Ribeiro et al. 2009). It plays multiple roles in the life cycle of the virus and in particular, P chaperones nascent RNA-free N (N^0), preventing its binding to cellular RNA, and deliver N^0 to the site of viral replication for the encapsidation of the newly replicated genomic RNA molecule (Mavrakis, Mehouas et al. 2006). The binding site for N^0 is situated in the first 60 N-terminal residues of P (Mavrakis, Mehouas et al. 2006).

In a recent study, we found that P proteins from different viruses share a common modular organization, and that the N-terminal region of RV P behaves as a disordered polypeptide as judged by its hydrodynamic radius measured by size exclusion chromatography (SEC) and its circular dichroism spectrum (Gérard, Ribeiro et al. 2009). These results contrast with the α-helical potential predicted for this region using bioinformatics approaches (Gérard, Ribeiro et al. 2009) and suggested that P_{NTD} could fold upon binding to N^0. Recently, we investigated the structural properties of the P_{NTD} region from VSV using a combination of Small Angle X-Ray Scattering (SAXS), NMR spectroscopy and Circular Dichroism (CD) experiments and demonstrated that VSV P_{NTD} possesses transiently populated α-helices and exists in solution as an ensemble of compacted and extended conformers. The polypeptide was shown to adopt more extended conformations in the presence of denaturants and to become more compact upon addition of stabilizing agents such as TMAO (Leyrat et al., *submitted, ARTICLE II*). Additionally, the secondary structure content estimated from the structural ensemble selected from the SAXS profile using the EOM approach was in good agreement with the values derived from both NMR and CD measurements, indicating that the combination of the EOM analysis with an atomic description of structural disorder was able to probe for secondary structure change when using the high Q range ($Q > 2.5$ nm^{-1}) of the SAXS profile, as stated by the theory (for a good review, see (Putnam, Hammel et al. 2007)).

Intrinsically disordered proteins (IDPs) are abundant in proteins from viruses and from higher eukaryotes. They are involved in many highly regulated processes such as viral pathogenesis, host cell interactions, cell cycle regulation, intracellular signaling, translation and transcription (Tokuriki, Oldfield et al. 2009). In particular, the increased plasticity of IDPs allows to couple a high specificity with a low affinity, enables binding of numerous structurally distinct targets, speeds up binding by a fly-casting mechanism, and provides the ability to overcome steric restrictions, enabling larger interaction surfaces in protein-protein and protein-ligand complexes than those obtained with rigid partners (Shoemaker, Portman et al. 2000; Dyson and Wright 2002; Dyson and Wright 2005; Tompa, Szasz et al. 2005; Uversky, Oldfield et al. 2005; Tompa and Fuxreiter 2008).

In this study, we investigate the folding potential of the N-terminal region of RV P comprising residues 1-68 (P_{68}) by varying a range of experimental conditions including protein concentration, denaturant Gdn-HCl, stabilizing co-solute TMAO, pH and temperature. From the analysis of SAXS data and CD, we show that P_{68} possesses some degree of structural pre-organization in solution and that its compactness and its helical content are highly sensitive to its environment. The effects of chaotropic and kosmotropic agents are compared with the data available for VSV P_{NTD} (Leyrat et al., *submitted, ARTICLE II*). The temperature-induced structural changes are analyzed in terms of α-helical and polyproline II content. Polyproline type II structure (PPII) has been described as an important extended conformational state populated by denatured or natively unfolded proteins and that tend to exist only in very short stretches of residues (Mezei, Fleming et al. 2004; Jha, Colubri et al. 2005; Zagrovic, Lipfert et al. 2005). The occurrence and the influence of experimental parameters on PPII content in IDRs and denatured proteins is a matter of considerable debate in the literature (Pappu and Rose 2002; Jha, Colubri et al. 2005; Whittington, Chellgren et al. 2005; Zagrovic, Lipfert et al. 2005; Kjaergaard, Norholm et al. 2010). The results presented for RV P_{68} based on SAXS data and ensemble optimization analysis allow to probe for the shape and secondary structure content of the disordered ensembles populate by P_{68} and thus provide valuable insights into the structural properties of this IDP, suggesting that P_{68} acts as a MoRE that fold upon binding the RNA free nucleoprotein N^0.

Buffer	pH	T(°C)	C (mg/ml)	R_g(nm)
50mM RE	7.5	20	3.3	2.78
-	-	-	6.6	2.58
-	-	-	33.3	2.12
TMAO 1M	-	-	3.3	2.56
Gdn-HCl 6M	-	-	6.6	2.81
50mM Glu, Arg	5.0	-	-	2.75
-	4.0	-	-	>5
-	7.5	5	3.9	2.70
-	-	20	-	2.66
-	-	40	-	2.62
-	-	60	-	2.55

Table 1: Rg values extracted from SAXS data

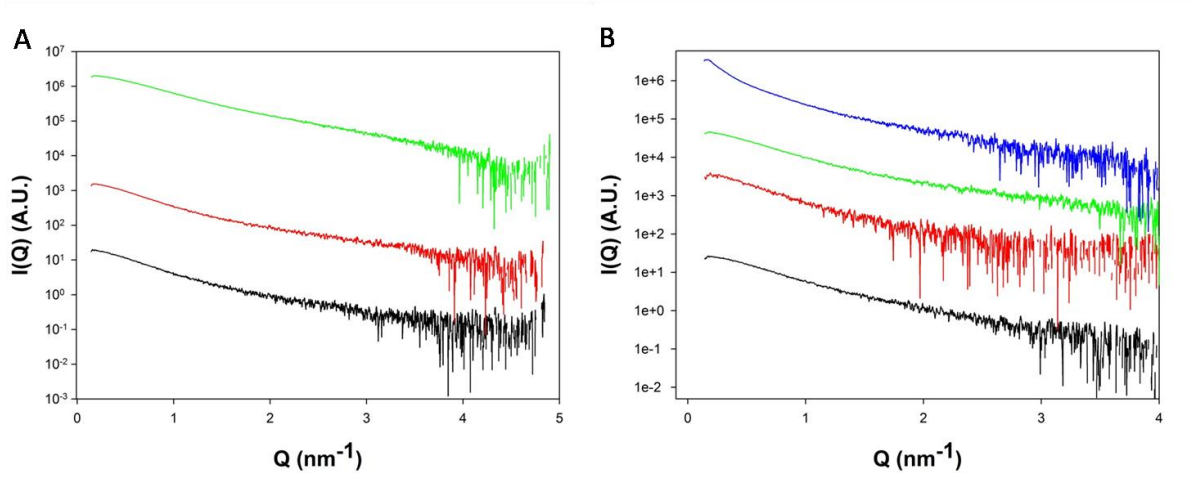


Figure 1: SAXS profiles of Rabies Virus P_{68}. A. SAXS data were recorded for Q values in the range 0.14 nm^{-1}< Q < 4.8nm^{-1}. The curves obtained for 3 protein concentrations (3mg/ml, 6mg/ml, and 33mg/ml) are represented in black, red, and green lines, respectively. **B.** SAXS profiles were recorded in the presence of 1M TMAO (black line), 6M Gdn-HCl (red line), 20mM Tris-HCl pH 5.0 (green line), or 20mM Tris-HCl pH 4.0 (blue line).

Results

P_{68} is able to compact and fold in a concentration dependent manner

We investigated the effect of protein concentration on the SAXS profile of P_{68} by varying the concentration between 3.3 and 33.3 mg/ml (Figure1A). The quality of the data is good up to Q ~ 3.5nm^{-1} and show no aggregation. The R_g values extracted from the data using Guinier approximation reveal a continuous decrease in the R_g of the molecule upon increasing protein concentration (Table 1). P_{68} compacts from R_g = 2.78 nm at 3.3 mg/ml to R_g = 2.12 nm at 33.3 mg/ml. RV P_{68} has been previously shown to display low (<5%) α-helical content in diluted solution, and its measured Stokes radius corresponds to that of a fully unfolded polypeptide (Gérard, Ribeiro et al. 2009). However, consensus sequence-based secondary structure predictions suggest the existence of α-helices between residues 10-38 and 51-56 (Gérard, Ribeiro et al. 2009). As RV P_{68} is disordered in solution and cannot be represented by a unique structure, we used an ensemble of 4000 atomic models to fit SAXS data using the EOM approach. The ensemble of randomly generated conformers is designed to include α-helical segments based on the predicted α-helical propensity of P_{68} (Figure 2A). The R_g distributions presented in Figure 2B suggests that the increase in concentration causes a depletion of extended conformers which is accompanied at higher concentration by a massive contraction of the molecule. The details of model selection indicate a gradual increase in the proportion of helical conformers (Figure2 C, D, E) which culminates at c = 33.3 mg/ml where nearly all the selected conformers bear a fully formed helix 1 (Figure 2E), suggesting that P_{68} undergoes a concentration-induced folding. We then calculated the PPII content in the initial and selected ensembles by counting the number of residues populating the Ramachandran region defined as φ = -75 ± 15° and ψ = 145 ± 15° and averaging over each ensemble. PPII content analysis shows that P_{68} is enriched in PPII at low concentration (14.4% at 3.3 mg/ml versus 10.1% for the initial ensemble), and that the selected populations of conformers tend to lose PPII content at high concentration to reach a value of 7.2% at 33.3 mg/ml. These observations are consistent with the overall compaction observed and indicate, together with the gain in α-helical conformers, a tendency of P_{68} to fold at high protein concentration.

P_{68} is extended in the presence of 6M Gdn-HCl but slightly compacts upon addition of 1M TMAO .

Next, we looked at the effect of denaturing- and stabilizing agents on the solution structure of P_{68} (Figure 1B, red and black lines, respectively). Because of the observed concentration-dependent folding of P_{68}, we chose to work at relatively low sample concentration (3.3 or 6.6 mg/ml) to minimize interparticle interference. Figure 3A shows the effect of adding 6M of Gdn-HCl on the R_g distribution of P_{68} for c = 6.6 mg/ml (data for c = 3.3 mg/ml were too noisy to be analyzed) and a comparison with the data recorded at the same protein concentration in 50 mM Glu, 50 mM Arg buffer. The denaturant causes a significant increase of the population of extended conformers with a concomitant increase of the average R_g (2.81 nm versus 2.58 nm). The observed R_g value corresponds closely to that expected for a random coil of this size (Fitzkee and Rose 2004; Bernado and Blackledge 2009). The model selection

histograms show that this increase in size is associated with a loss of the existing fraction of

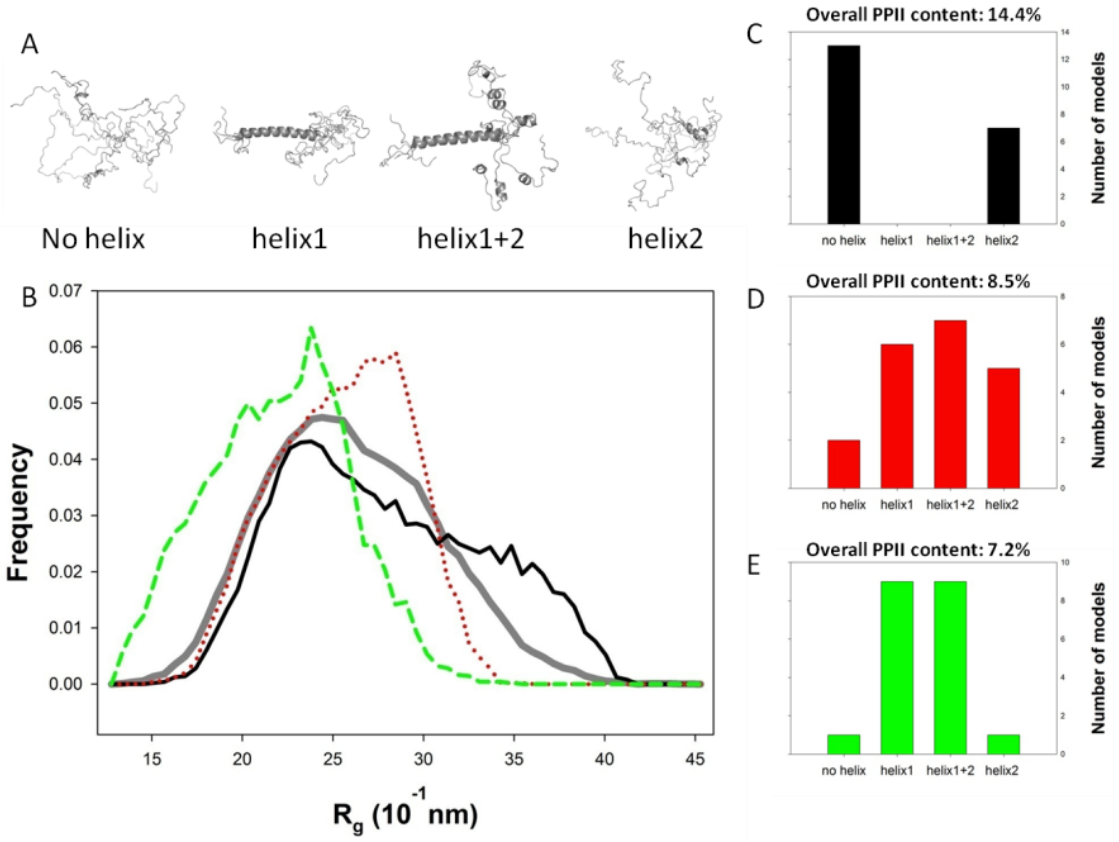


Figure 2: Effect of sample concentration on P_{68} structure. A. Models used in the EOM procedure. 1000 models were generated using FlexibleMecchano (Bernado, Blanchard et al. 2005) by including either no helical structure (no helix), one helix between residues 10 to 38 (helix1), 2 helices between residues 10 to 38 and 51 to 56 (helix1+2), or one short helix between residues 51 to 56 (helix2) according to secondary structure predictions (Gérard, Ribeiro et al. 2009). **B. EOM analysis.** The grey curve shows the Rg distribution calculated for the initial ensemble of conformers, whereas the black,red and green curves show the Rg distribution of the selected ensemble that fits the experimental SAXS data recorded at 3mg/ml, 6mg/ml or 33mg/ml, respectively. **C, D, E. Selected Models.** The histogram bars represent the number of models from each type (described in A.) that were selected by the genetic algorithm for the 3 different concentrations of P_{68} (corresponding to the Rg distributions shown in B). The average content of PPII helices in the ensemble is shown above each histograms.

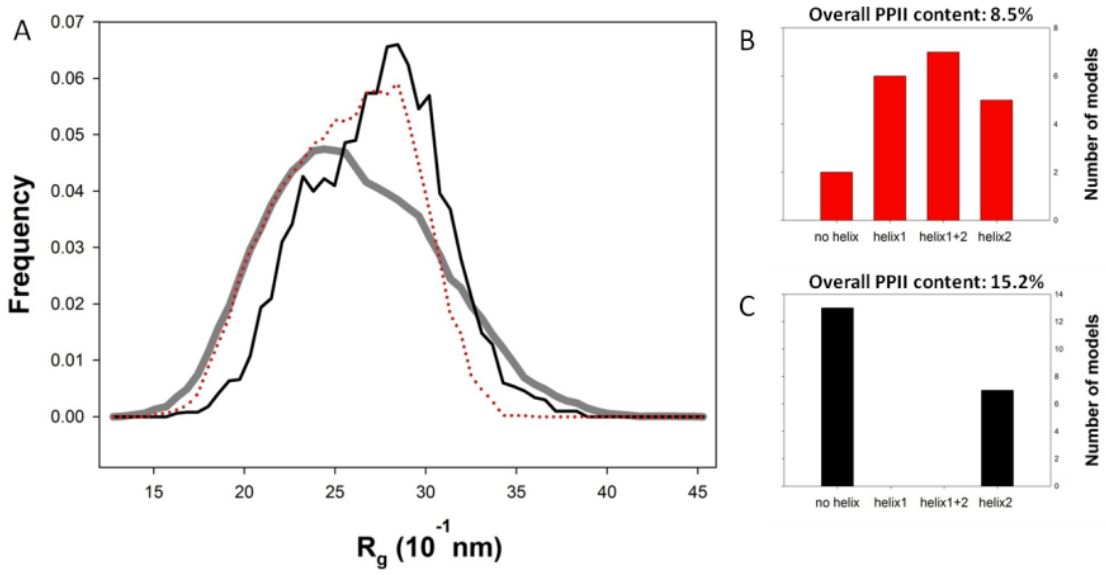


Figure 3: Effect of Guanidinium Hydrochloride (Gdn-HCl) on P_{68} structure. A. EOM analysis. The grey curve shows the Rg distribution calculated for the initial ensemble of conformers, whereas the black and red curves show the Rg distribution of the selected ensemble that fits the experimental SAXS data recorded 6mg/ml in the presence of Gdn-HCl 6M or in 50mM Glu,Arg buffer pH 7.5,

respectively. **B and C. Selected Models.** The histogram bars represent the number of models from each type (described in Figure 2A.) that were selected by the genetic algorithm in the presence or in the absence of 6M Gdn-HCl.

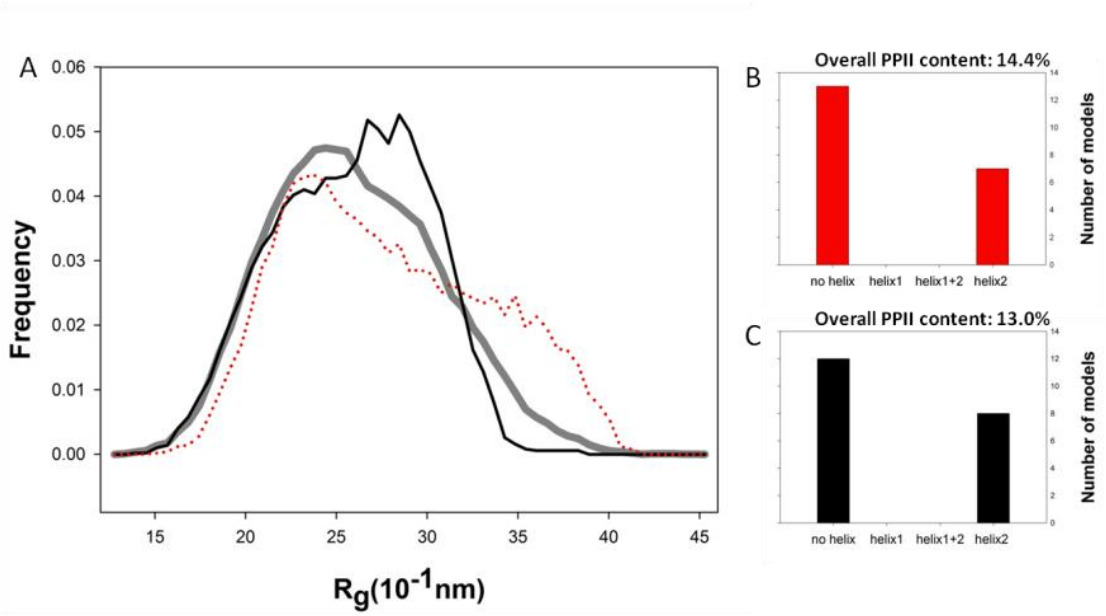


Figure 4: Effect of Trimethyl Amine N-Oxyde (TMAO) on P_{68} structure. A. EOM analysis. The grey curve shows the Rg distribution calculated for the initial ensemble of conformers, whereas the black and red curves show the Rg distribution of the selected ensemble that fits the experimental SAXS data recorded 3mg/ml in the presence of TMAO at 1M or in 50mM Glu,Arg buffer pH 7.5, respectively. **B and C. Selected Models.** The histogram bars represent the number of models from each type (described in Figure 2A.) that were selected by the genetic algorithm in the presence or in the absence of 1M TMAO.

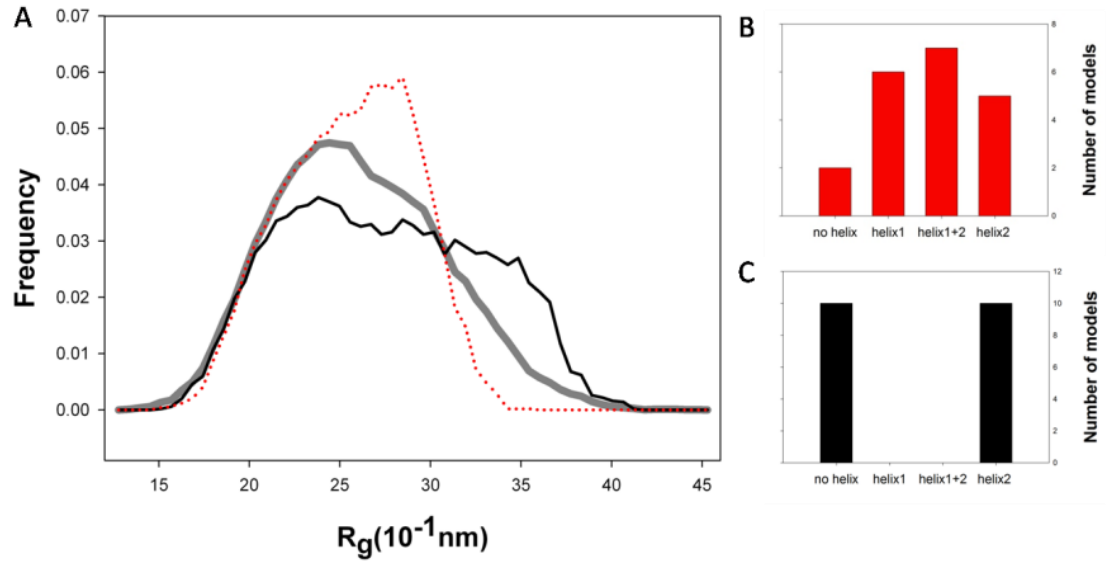
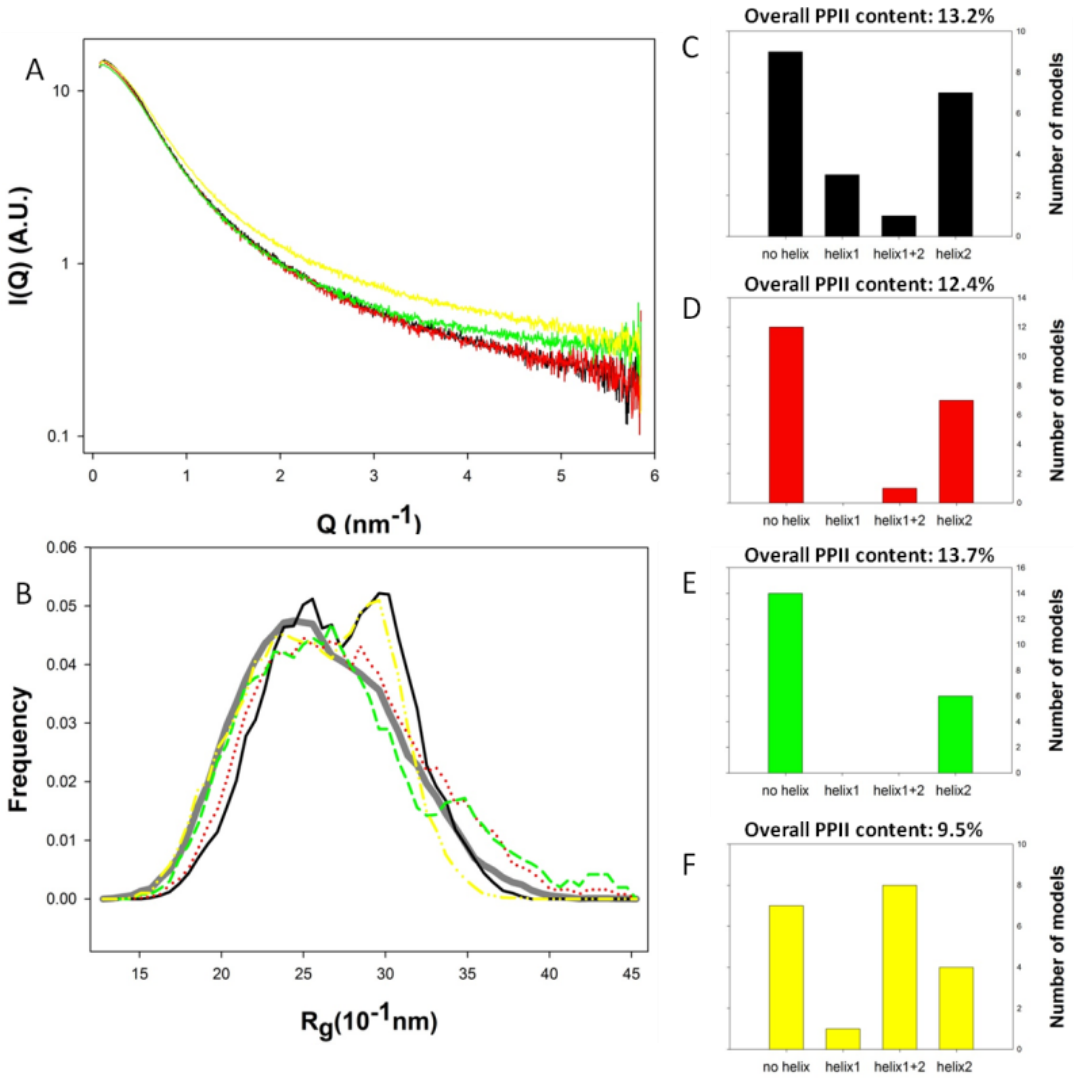


Figure 5: Effect of decreasing pH on P_{68} structure. A. EOM analysis. The grey curve shows the Rg distribution calculated for the initial ensemble of conformers, whereas the black and red curves show the Rg distribution of the selected ensemble that fits the experimental SAXS data recorded at 6mg/ml in 50mM Glu,Arg buffer at pH 5.0 or at pH 7.5, respectively. **B and C. Selected Models.** The histogram bars represent the number of models from each type (described in Figure 2A.) that were selected by the genetic algorithm at pH 5.0 or at pH 7.5.

conformers bearing helix1, leaving only models with no or only very short α-helical fragments in the selected ensemble. Concomitantly, the proportion of PPII conformations in the ensemble changes from 8.5% to 15.2%, suggesting that the denaturant favors PPII conformers. The addition of 1M TMAO to P_{68} has an opposite effect on the SAXS profile. Indeed, the addition of this stabilizing agent leads to a small but significant decrease in the calculated R_g (2.56 nm versus

2.78nm at an identical concentration of 3.3 mg/ml) (Table 1) which is associated with a decrease in the population of extended conformers (Figure 4A).



However, this compaction was not

Figure 6: Effect of temperature on P_{68} structure. A. SAXS profiles of P_{68} at different temperatures. SAXS data were recorded for Q values in the range 0.14 nm^{-1}< Q < 5.9nm^{-1} with a fixed sample concentration of 3.9 mg/ml in 50mM Glu, Arg buffer pH 7.5. Temperature was varied between 5°C and 60°C. **B. EOM analysis.** The grey curve shows the Rg distribution calculated for the initial ensemble of conformers, whereas the black, red, green, and yellow curves show the Rg distribution of the selected ensemble that fits the experimental SAXS data recorded at 5, 20, 40 and 60°C, respectively. **C, D, E and F. Selected Models.** The histogram bars represent the number of models from each type (described in Figure 2A.) that were selected by the genetic algorithm for the 4 different temperatures at which P_{68} was recorded.

correlated with any detectable increase in the selection of helical conformers (Figure 4B and C), and only a slight reduction of the PPII content was observed which suggest that TMAO has only a mild effect on the structure of P_{68}.

P_{68} structure is stabilized at low pH and undergoes increased interparticle interactions and aggregation .

The effect of acidic pH on P_{68} structure was studied by recording SAXS profiles and CD spectra at various pH values ranging from 4.0 to 8.5 (Figures 5 and 7). The R_g distribution at pH 8.5 is essentially identical to the data obtained at pH 7.5 (data not shown). However, dialysis of P_{68} against 20mM Tris-HCl buffer at pH 4.0 caused the formation of soluble aggregates as shown by its SAXS profile and measured R_g > 5nm (Figure 1 and Table 1). At pH 5.0, however, the measured R_g is close to 2.75nm, and the EOM R_g distribution shows an increase in the population of large conformers (Figure 5A). This increase in size is linked with a loss of helical conformers which is in

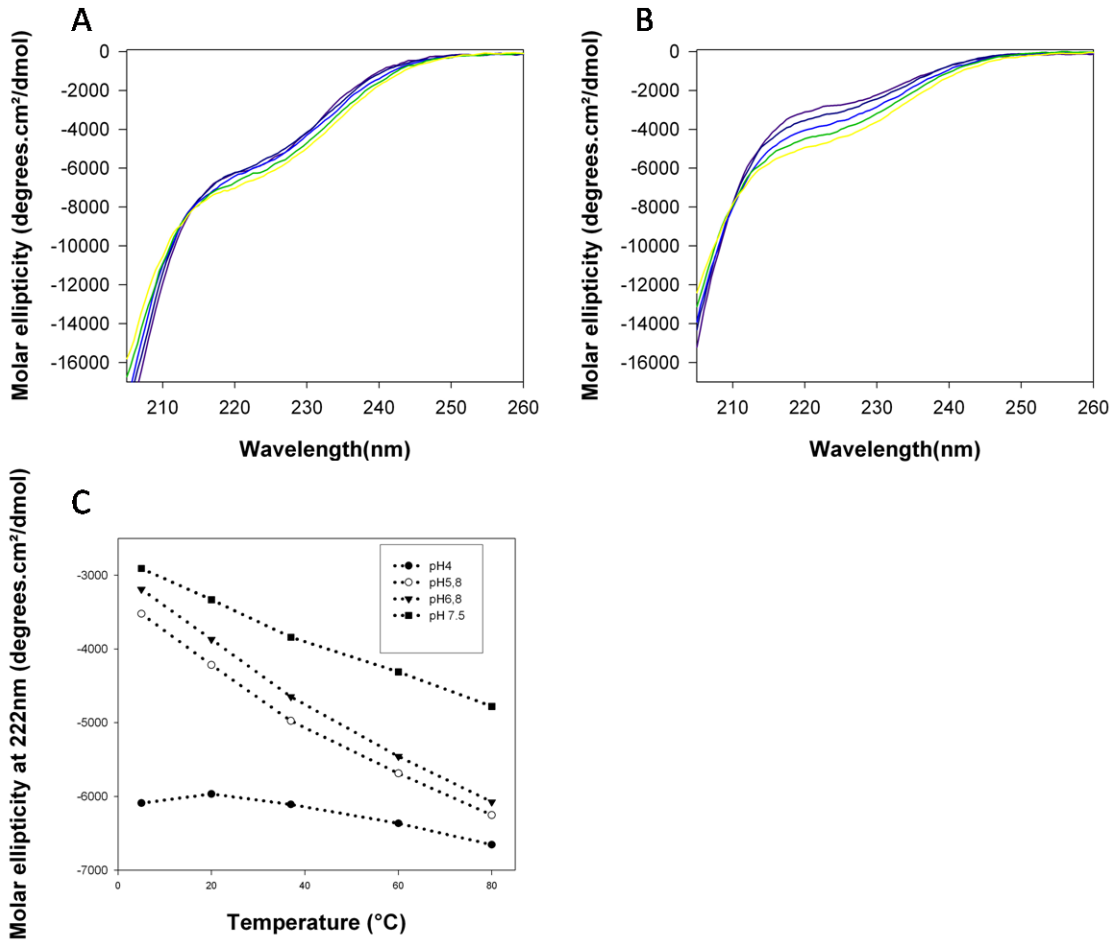


Figure 7: Effect of temperature and pH on the Far-UV Circular Dichroism spectrum of P_{68}. A. Effect of temperature on the CD spectrum of P_{68} at pH 7.5 between 5 and 80°C (from dark blue to yellow). **B.** Effect of temperature on the CD spectrum of P_{68} at pH 4.0 between 5 and 80°C (from dark blue to yellow). **C.** Molar ellipticity at 222nm is plotted versus temperature for P_{68} at pH 4.0, 5.8, 6.8 and 7.5.

apparent contradiction with the slight increase in molar ellipticity at 222 nm measured by CD (Figure 7).

Temperature induces P_{68} compaction and affects its helical propensity

Temperature dependent conformational changes have been studied with many IDPs (Uversky, Li et al. 2001; Dawson, Muller et al. 2003; Gast, Zirwer et al. 2003; Jarvet, Damberg et al. 2003; Sanchez-Puig, Veprintsev et al. 2005; Malm, Jonsson et al. 2007; Jeganathan, von Bergen et al. 2008; Uversky 2009). Recently, Kjaergaard and co-workers showed by using chemical shift analysis with intrinsic referencing based on NMR measurements performed between 5° and 45°C on several disordered peptides that the structural change observed upon heating was dominated by a loss of PPII conformations and was accompanied by a loss of α-helices (Kjaergaard, Norholm et al. 2010). Here, we recorded SAXS profiles of P_{68} between 5 and 60°C (Figure 6A). We observe a general compaction effect with increasing temperature as shown by the measured R_g values (Table 1), a phenomenon that was previously observed both for natively unfolded and denatured proteins (Noppert, Gast et al. 1996; Gast, Zirwer et al. 1997; Sadqi, Lapidus et al. 2003; Nettels, Muller-Spath et al. 2009). The change is also apparent on the curve in the form of reduced intensity at low Q values and a higher signal at high Q values, especially at high (60°C) temperature, which likely reflects changes in P_{68} structure (Figure 6A). The EOM profile confirms this tendency and shows a gradual compaction of the structural ensemble (Figure 6B). At intermediate temperatures (20°C and 40°C), the structural ensemble is slightly flattened, and includes both more compact and more extended structures than at 5°C, which is in agreement with increased disorder of the peptide chain and loss of secondary structure that was observed with various disordered peptides using NMR (Kjaergaard, Norholm et al. 2010). Less conformers containing α-helical segments are selected at 40°Cthan at 5°C (Figure 6C, D, E). However, at 60°C, the proportion of α-helical conformers increases significantly (Figure 6F) while the population of

extended conformers decreases (Figure 6B). The average PPII content calculated for the selected ensembles displays only minor fluctuations between 5°C and 40°C but is markedly reduced at 60°C and drops from 13.7% to 9.5%.

In order to confirm these results, the temperature dependence of the CD spectra of P_{68} was recorded at different pH values ranging from 4 to 7.5 (Figure 7). The spectra display a decreasing shoulder at 222 nm and an increasing signal in the 205-210 nm region when the temperature is increased between 5° and 80°C, which can be interpreted either in terms of a gain in α-helical structure or by a loss of PPII conformations. The changes observed were reversible up to 85°C (data not shown). However, as can be seen in Figure 7C, at pH values higher than 5, the decrease in CD signal at 222nm is not linear, suggesting that different processes occur. At pH 4, the signal seems not to be affected by temperature when T is comprised between 5 and 40°C and significantly drops only above 40°C. The difference in the temperature dependence of the CD signal at high and low pH suggests that temperature exerts at least two distinct effects on P_{68} structure, and that the effect observed at moderately high temperature ($5 < T < 40$°C) is pH dependent.

Discussion

Solution structure of the N-terminal region of P

The data presented here suggests that RV P_{68} can be adequately represented as a structural ensemble of conformers of different compactness with transient helical structures. The SAXS technique is known to be sensitive to protein shape at medium resolution (1-3 nm^{-1}) and to small structural fluctuations at high resolution (> 3 nm^{-1}) (for a review, see (Putnam, Hammel et al. 2007)), and structural information from SAXS has been used to guide the atomic-scale modeling of proteins and their assemblies (Zheng and Doniach 2002; Wu, Tian et al. 2005; Zheng and Doniach 2005; Förster, Webb et al. 2008). EOM analysis of SAXS data recorded at different concentrations suggested that the compaction observed at high concentrations results in a population shift towards conformers containing in average more α-helical segments and less PPII. This effect can be explained in terms of repulsive intermolecular interactions arising from the high net charge of P_{68} (-15 at neutral pH), which causes the protein to compact and stabilizes the α-helical conformers at high concentration, while disfavoring the extended PPII helices. We recently reported on the EOM analysis of VSV P_{NTD} SAXS data (Leyrat et al., 2010); interestingly, the concentration dependent compaction of the molecule was not observed in this case, presumably due to the lower net charge (-10) of the N–terminal region of VSV P. The N-terminal region of P had been previously shown to interact with N in the N^0-P complex and to be required for encapsidation of the viral RNA (Mavrakis, Mehouas et al. 2006; Chen, Ogino et al. 2007). This complex forms in the cytoplasm of infected cells and maintains nascent N molecules in a soluble and RNA-free form that is competent for encapsidation of the newly synthesized RNA genomes. In both VSV and RV, the RNA binding cavity of N has a high electrostatic surface potential (Albertini, Wernimont et al. 2006; Green, Zhang et al. 2006), whereas P_{NTD} is rich in negatively charged residues, suggesting that P_{NTD} binds in the same cavity as RNA (Mavrakis, Mehouas et al. 2006).

Recently, we developed a method that combines multiple predictions of disorder to localize the borders between disordered and structured regions in a protein from its amino acid sequence (Ribeiro, Favier et al. 2008; Gérard, Ribeiro et al. 2009). This approach predicted the presence of a structured region at the N-terminal extremity of both RV and VSV P proteins. However, various experimental results indicate that the isolated N-terminal region of RV ((Gérard, Ribeiro et al. 2009) and this study) and VSV P (Leyrat et al., 2010) are highly flexible. The chemical shift dispersion and line width in the NMR spectra of VSV P_{NTD} are typical of unfolded proteins, and for both RV and VSV, the values of the hydrodynamic radius measured by SEC and of the radius of gyration measured by SAXS are close to those expected for unfolded proteins. By contrast, CD and NMR dynamics revealed the presence of small amounts of fluctuating helical structures in two regions of VSV P_{NTD}. In addition, the modeling of VSV P_{NTD} as ensembles of conformers on the basis of SAXS data revealed the co-existence of two sub-populations with different degrees of compactness. For RV P_{68}, the bimodal nature of the R_g distribution is apparent at low concentration and an increased occurrence of compact conformers is induced at higher concentration. This suggests the existence of a sub-population of compact conformers that could prefigure the conformation adopted by this protein region in the complex formed with its natural partner, the RNA-free nucleoprotein.

Effects of co-solvents on P_{68} ensembles

The changes in compactness observed in the presence of Gdn-HCl and TMAO for RV P_{68} are comparable to the results obtained for VSV. The addition of denaturant caused a decrease in the population of compacted conformers and the disappearance of partially folded conformers in the selected ensemble. This is consistent with the observation that denatured proteins tend to have slightly larger radii than natively unfolded proteins

(Uversky 2002). Moreover, the proportion of PPII conformations increases significantly in the presence of denaturant Gdn-HCl, an effect which has been previously reported for urea-unfolded proteins (Whittington, Chellgren et al. 2005; Pace, Huyghues-Despointes et al. 2010). On the contrary, the addition of 1M TMAO led to the depletion of the most extended conformers through its solvophobic effect. TMAO has been suggested to favor protein folding through destabilization of protein backbone interactions with water (Baskakov and Bolen 1998; Baskakov, Kumar et al. 1999; Uversky, Li et al. 2001; Mello and Barrick 2003). Interestingly, TMAO has a milder effect on the folding of RV P_{NTD} compared to VSV, for which a higher proportion of α-helical conformers was induced in the presence of TMAO (30 % of α-helical content). This difference might arise from the lower α-helical potential of RV compared to VSV (Gérard, Ribeiro et al. 2009), a feature that might, in turn, be related to the stronger electrostatic repulsions in RV P_{NTD}.

Effects of pH and temperature on P_{68} ensembles

The effect of lowering the pH on acidic IDPs has been previously correlated with an increased hydrophobic force resulting from the decrease in the net charge of the polypeptide (Uversky, Li et al. 2001). In the case of RV P_{68}, acidification is associated with increased secondary structure content as measured by CD. However, the decrease in repulsive interactions causes the protein to aggregate at the higher concentrations used in SAXS experiments leading to an increase in the measured R_g at pH 5.0 and to major aggregation at pH 4.0.

A characteristic temperature-dependent spectroscopic change has been observed in numerous IDPs (Uversky, Li et al. 2001; Dawson, Muller et al. 2003; Gast, Zirwer et al. 2003; Jarvet, Damberg et al. 2003; Sanchez-Puig, Veprintsev et al. 2005; Malm, Jonsson et al. 2007; Jeganathan, von Bergen et al. 2008; Uversky 2009; Kjaergaard, Norholm et al. 2010). The structural change that can be recorded by CD could be interpreted either in terms of loss of PPII conformations or by the formation of α-helices. Recently, a combined CD, SAXS and NMR study found that the structural change was likely to result from a loss of PPII conformations. The three studied IDPs were found to lose rather than gain α-helices when comparing NMR measurements performed at 5° and 45°C, whereas a decrease in the measured R_g was also observed by SAXS (Kjaergaard, Norholm et al. 2010). However, the integrality of the spectroscopic change could not be followed neither by NMR nor SAXS at temperatures higher than 45°C, and the angular range of the SAXS data was limited to 2.5 nm^{-1}, which is insufficient to probe the average secondary structure content of IDPs. Here, SAXS profiles with acceptable signal-to-noise ratios were fitted up to $Q = 5\ nm^{-1}$ for temperatures ranging from 5° to 60°C. The curves reveal significant changes in the high Q region with increasing temperature. The structural ensembles selected by EOM are almost totally depleted in α-helical conformers as the temperature is raised from 5 to 40°C. This is correlated with broadening of the R_g distribution of the selected ensemble and a slight R_g decrease. This effect might reflect a temperature-induced redistribution of the random coil associated with a loss of PPII conformations as suggested previously (Kjaergaard et al., 2010). However, the profile obtained at 60°C contrasts with the previous observations and shows an increased occurrence of α-helical conformers and a more pronounced compaction effect, which is accompanied by significant loss of extended PPII conformations. This temperature-induced change is clearly visible in the SAXS pattern of 60°C-heated P_{68} and changes in the high Q region of the SAXS curve can be theoretically associated with changes in secondary structure content (Putnam, Hammel et al. 2007). Alternatively, a temperature-induced stabilization of α-helical conformers through an increase in hydrophobic effect, as suggested by Uversky and colleagues (Uversky, Li et al. 2001), might be apparent only at moderately high temperature and be masked below 40-45°C by statistical coil redistribution. Interestingly, the CD spectra recorded at different temperatures and pH values show a linear change in signal which can be divided in two regions, as the slope slightly changes above 40°C. This is shown more clearly at lower pH (4.0) where the signal is nearly constant between 5 and 40° and does only decrease between 40 and 80°C, revealing the existence of two regions, consistently with our interpretation derived from ensemble analysis of SAXS data. The effect of temperature on IDPs is complex in nature and seems to involve simultaneous changes in conformational entropy, hydrophobic effect and intra-molecular hydrogen bonding (Uversky, Li et al. 2001; Nettels, Muller-Spath et al. 2009; Kjaergaard, Norholm et al. 2010). Clearly, much more work is needed to fully elucidate the mechanisms of temperature-induced structural changes in IDPs.

Implications for recognition of N^0 by the N-terminal region of P

Our data shows that P_{68} is highly flexible and has average molecular dimensions of a random coil chain, but contains fluctuating α-helical structures, which together with the existence of a compact sub-population, indicates some degree of pre-organization in the N-terminal region of P. The apparent discrepancy between bioinformatics predictions (Gérard, Ribeiro et al. 2009) and the

experimental observations of disorder suggests that this region of P could fold upon binding to its viral partner, the RNA-free N^0. Because residues 4 to 60 of P have been implicated in binding to N^0 (Mavrakis, Mehouas et al. 2006), the existence of pre-organized conformers in solution suggests that it has the potential to fold but has not enough intra-molecular interactions to exist by itself in a folded conformation. Binding to its partner would provide additional inter-molecular interactions that could stabilize the structure and lower its intramolecular repulsion caused by its high net charge.

In the viral replication cycle, the N^0-P complex forms only transiently as an intermediate during the synthesis of new nucleocapsids. The N^0-bound P must be outcompeted by the newly synthesized genomic RNA, and therefore the affinity of P for N^0 cannot be too high. The mechanism of folding upon binding has been proposed as a solution for providing a highly specific recognition of a ligand by its receptor, without the corollary of a high affinity (Dyson and Wright 2002; Wright and Dyson 2009). Generally, complexes formed between a disordered protein and a structured one, have large binding interfaces (Mészáros, Tompa et al. 2007). The folding or the adoption of a rigid structure by the ligand upon binding to its receptor leads to the formation of multiple specific intermolecular interactions that provide a great specificity. A large number of contacts would generally results in an extremely high affinity, but in the case of a ligand that is disordered in its unbound form, this strong binding energy is opposed by the high entropy of the disordered protein. Such a mechanism could explain how the N-terminal region of P specifically recognizes the RNA-free N^0 molecule, in particular through electrostatic complementation, but then could be displaced by the viral RNA. The mechanism of folding upon binding to a partner remain poorly understood (Wright and Dyson 2009), and two models have been proposed that involved either an induced fit mechanism in which binding is coupled with folding or a selection mechanism in which preexisting equilibrium conformations are selected (Tsai, Kumar et al. 1999). Both the presence of fluctuating α-helical structures in P_{68} and the co-existence of sub-populations of compact and of extended conformers support the existence of preformed structural elements. However, in the structure of the N-RNA complex (Albertini, Wernimont et al. 2006), the gap between the N- and C-terminal domains of N is small, suggesting that a folded, compact form of P_{NTD} could not insert into the RNA-binding grove without a significant conformational change in N. This structural feature of N rather support a model of coupled folding and binding in which P_{NTD} would bind to N^0 in a relatively extended conformation through non native interactions and electrostatic attractions and subsequently folds within the N^0-P complex, as has been suggested for other complexes involving a disorder-to-order transition (Sugase, Dyson et al. 2007; Turjanski, Gutkind et al. 2008). It is now important to obtain structural information about the structure of the N^0-P complex in order to understand how P_{NTD} interacts with N^0 and how N^0 is transferred onto newly synthesized RNA genomes.

Material and Methods

Protein sample preparation

Rabies Virus phosphoprotein residues 1-68 was produced and purified according to previously published procedure (Gérard, Ribeiro et al. 2009). Sample quality was checked using size exclusion chromatography (SEC) combined with detection by multi-angle laser light scattering (MALLS) and refractometry (RI) (Gérard, Ribeiro et al. 2007; Hodak, Wohlkönig et al. 2008).

Circular dichroism spectroscopy

Far-UV circular dichroism spectra were recorded at 20°C on a JASCO model J-810 CD spectropolarimeter equipped with a Peltier temperature controller. RV P_{NTD} was dialysed against 20mM Tris-HCl at appropriate pH (4-8.5) and diluted to final concentrations of about 40 µM in dH_2O. Spectra were measured in a cuvette with a path length of 2 mm. After subtracting the blank signal, the CD signal (in millidegrees) was converted to mean molar residue ellipticity (in $deg.cm^2.dmol^{-1}$).

Small-angle X-ray scattering (SAXS) experiments

The monodispersity of the samples used in SAXS experiments was checked by size exclusion chromatography (SEC) combined with detection by multi-angle laser light scattering (MALLS) and refractometry (RI) (Gérard, Ribeiro et al. 2007; Hodak, Wohlkönig et al. 2008). SAXS data were collected at the European Synchrotron Radiation Facility (Grenoble, France) on beamline ID14-3. The sample-to-detector distance was 1 m and the wavelength of the X-rays was 0.0995 nm. Samples were contained in a 1.9 mm wide quartz capillary. The time of exposure was optimized for reducing radiation damage. Data acquisition was performed at 20 °C. Protein concentrations ranged from 3 to 33 $mg.ml^{-1}$. Data reduction was performed using the established procedure available at ID14-3, and buffer background runs were subtracted from sample runs. The radius of gyration and forward intensity at zero angle (*I(0)*) were determined with the programs PRIMUS (Konarev, Volkov et al. 2003) by using the Guinier approximation at low *Q* values, in a $Q.R_g$ range up to 1.3:

$$\ln I(Q) = \ln I(0) - \frac{R_g^2 Q^2}{3}$$

The forward scattering intensity was calibrated using bovine serum albumin and lysozyme as references. The radius of gyration and pairwise distance distribution function, *P(r)*, were calculated with the program GNOM (Semenyuk and Svergun 1991). The maximum dimension (D_{max}) value was adjusted so that the R_g value obtained from GNOM agreed with that obtained from the Guinier analysis.

Structural Modeling

A conformational ensemble of 4,000 conformers was generated with the program Flexible-Meccano (Bernado, Blanchard et al. 2005) by pooling 4 sub-ensembles of 1000

conformers in which different amounts of helical structure were imposed in the putative helical regions of P_{NTD} identified by secondary structure prediction (Gérard, Ribeiro et al. 2009) as follows: Ensemble 1: no helix; Ensemble 2: aa 10-38; Ensemble 3: aa 10-38 and aa 51-56; Ensemble 4: aa 51-56. An optimized ensemble of conformations that agrees with the experimental SAXS data was selected from the large conformational ensemble using the Ensemble Optimization Method (EOM) software (Bernado, Mylonas et al. 2007). In order to calculate polyproline II helix (PPII) content, the pool ensemble and all the selected ensembles were converted into trajectories using Gromacs (Hess, Kutzner et al. 2008), and processed using Gromacs tool g_rama to calculate Ramachandran plots, from which the PPII content of each ensemble was extracted (defined by $\varphi = -75 \pm 15°$ and $\psi = 145 \pm 15°$).

References

Albertini, A. A., A. K. Wernimont, et al. (2006). "Crystal structure of the rabies virus nucleoprotein-RNA complex." Science **313**(5785): 360-3.

Albertini, A. A. V., A. K. Wernimont, et al. (2006). "Crystal structure of the rabies virus nucleoprotein-RNA complex." Science **313**(5785): 360-360.

Arnheiter, H., N. L. Davis, et al. (1985). "Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis." Cell **41**(1): 259-67.

Baskakov, I. and D. W. Bolen (1998). "Forcing thermodynamically unfolded proteins to fold." J Biol Chem **273**(9): 4831-4.

Baskakov, I. V., R. Kumar, et al. (1999). "Trimethylamine N-oxide-induced cooperative folding of an intrinsically unfolded transcription-activating fragment of human glucocorticoid receptor." J Biol Chem **274**(16): 10693-6.

Bernado, P. and M. Blackledge (2009). "A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering." Biophys J **97**(10): 2839-45.

Bernado, P., L. Blanchard, et al. (2005). "A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering." Proc Natl Acad Sci U S A **102**(47): 17002-7.

Bernado, P., E. Mylonas, et al. (2007). "Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering." J Am Chem Soc.

Chen, M., T. Ogino, et al. (2007). "Interaction of vesicular stomatitis virus P and N proteins: Identification of two overlapping domains at the N-terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA." J Virol **81**: 13478-13485.

Dawson, R., L. Muller, et al. (2003). "The N-terminal domain of p53 is natively unfolded." J Mol Biol **332**(5): 1131-41.

Dyson, H. J. and P. E. Wright (2002). "Coupling of folding and binding for unstructured proteins." Curr Opin Struct Biol **12**(1): 54-60.

Dyson, H. J. and P. E. Wright (2005). "Intrinsically unstructured proteins and their functions." Nat Rev Mol Cell Biol **6**(3): 197-208.

Fields, B. N., D. M. Knipe, et al. (1996). Fields' virology, 3rd edn. New York, Lippincott-Raven Publishers.

Fitzkee, N. C. and G. D. Rose (2004). "Reassessing random-coil statistics in unfolded proteins." Proc Natl Acad Sci U S A **101**(34): 12497-502.

Förster, F., B. Webb, et al. (2008). "Integration of small-angle X-ray scattering data into structural modeling of proteins and their assemblies." Journal of Molecular Biology **382**(4): 1089-1106.

Gast, K., D. Zirwer, et al. (2003). "Are there temperature-dependent structural transitions in the "intrinsically unstructured" protein prothymosin alpha?" Eur Biophys J **31**(8): 586-94.

Gast, K., D. Zirwer, et al. (1997). "Ribonuclease T1 has different dimensions in the thermally and chemically denatured states: a dynamic light scattering study." FEBS Lett **403**(3): 245-8.

Gérard, F., E. Ribeiro, et al. (2007). "Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution." Biochemistry **46**: 10328-10338.

Gérard, F. C. A., E. A. Ribeiro , et al. (2009). "Modular organization of rabies virus phosphoprotein." J. Mol. Biol. **388**: 978-996.

Green, T. J., X. Zhang, et al. (2006). "Structure of the vesicular stomatitis virus nucleoprotein-RNA complex." Science **313**(5785): 357-60.

Hess, B., C. Kutzner, et al. (2008). "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation." J. Chem. Theory Comput. **4**: 435-447.

Hodak, H., A. Wohlkönig, et al. (2008). "The peptidyl-prolyl isomerase and chaperone Par27 of *Bordetella pertussis* as the prototype for a new group of parvulins." J Mol Biol **376**(2): 414-26.

Iseni, F., A. Barge, et al. (1998). "Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures." J Gen Virol **79 (Pt 12)**: 2909-19.

Jarvet, J., P. Damberg, et al. (2003). "A left-handed 31 helical conformation in the Alzheimer A [beta](12-28) peptide." FEBS letters **555**(2): 371-374.

Jeganathan, S., M. von Bergen, et al. (2008). "The natively unfolded character of tau and its aggregation to Alzheimer-like paired helical filaments." Biochemistry **47**(40): 10526-39.

Jha, A. K., A. Colubri, et al. (2005). "Statistical coil model of the unfolded state: resolving the reconciliation problem." Proc Natl Acad Sci U S A **102**(37): 13099-104.

Jha, A. K., A. Colubri, et al. (2005). "Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library." Biochemistry **44**(28): 9691-702.

Kjaergaard, M., A. B. Norholm, et al. (2010). "Temperature-dependent structural changes in intrinsically disordered proteins: formation of alpha-helices or loss of polyproline II?" Protein Sci **19**(8): 1555-64.

Konarev, P. V., V. V. Volkov, et al. (2003). "PRIMUS: a Windows PC-based system for small-angle scattering data analysis." J. Appl. Cryst. **36**: 1277-1282.

Malm, J., M. Jonsson, et al. (2007). "Structural properties of semenogelin I." FEBS J **274**(17): 4503-10.

Mavrakis, M., S. Mehouas, et al. (2006). "Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA." Virology **349**(2): 422-9.

Mello, C. C. and D. Barrick (2003). "Measuring the stability of partly folded proteins using TMAO." Protein Sci **12**(7): 1522-9.

Mészáros, B., M. Tompa, et al. (2007). "Molecular Principles of the Interactions of Disordered Proteins." J. Mol. Biol. **372**: 549-561.

Mezei, M., P. J. Fleming, et al. (2004). "Polyproline II helix is the preferred conformation for unfolded polyalanine in water." Proteins **55**(3): 502-7.

Nettels, D., S. Muller-Spath, et al. (2009). "Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins." Proc Natl Acad Sci U S A **106**(49): 20740-5.

Noppert, A., K. Gast, et al. (1996). "Reduced-denatured ribonuclease A is not in a compact state." FEBS Lett **380**(1-2): 179-82.

Pace, C. N., B. M. P. Huyghues-Despointes, et al. (2010). "Urea denatured state ensembles contain extensive secondary structure that is increased in hydrophobic proteins." Protein Science **19**(5): 929-943.

Pappu, R. V. and G. D. Rose (2002). "A simple model for polyproline II structure in unfolded states of alanine-based peptides." Protein Sci **11**(10): 2437-55.

Putnam, C. D., M. Hammel, et al. (2007). "X-ray solution scattering (SAXS) combined with crystallography

and computation: defining accurate macromolecular structures, conformations and assemblies in solution." Q Rev Biophys **40**(3): 191-285.

Ribeiro, E. A., Jr., A. Favier, et al. (2008). "Solution Structure of the C-Terminal Nucleoprotein-RNA Binding Domain of the Vesicular Stomatitis Virus Phosphoprotein." J Mol Biol **382**: 525-538.

Sadqi, M., L. J. Lapidus, et al. (2003). "How fast is protein hydrophobic collapse?" Proc Natl Acad Sci U S A **100**(21): 12117-22.

Sanchez-Puig, N., D. B. Veprintsev, et al. (2005). "Human full-length Securin is a natively unfolded protein." Protein Sci **14**(6): 1410-8.

Schoehn, G., F. Iseni, et al. (2001). "Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site." J Virol **75**(1): 490-8.

Semenyuk, A. V. and D. Svergun (1991). "GNOM - a program package for small-angle scattering data processing." J. Appl. Crystallog. **24**: 537-540.

Shoemaker, B. A., J. J. Portman, et al. (2000). "Speeding molecular recognition by using the folding funnel: the fly-casting mechanism." Proc Natl Acad Sci U S A **97**(16): 8868-73.

Sugase, K., H. J. Dyson, et al. (2007). "Mechanism of coupled folding and binding of an intrinsically disordered protein." Nature **447**: 920-921.

Tokuriki, N., C. J. Oldfield, et al. (2009). "Do viral proteins possess unique biophysical features?" Trends Biochem Sci **34**(2): 53-9.

Tompa, P. and M. Fuxreiter (2008). "Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions." Trends Biochem Sci **33**(1): 2-8.

Tompa, P., C. Szasz, et al. (2005). "Structural disorder throws new light on moonlighting." Trends Biochem Sci **30**(9): 484-9.

Tsai, C. J., S. Kumar, et al. (1999). "Folding funnels, binding funnels, and protein function." Protein Sci **8**(6): 1181-90.

Turjanski, A. G., J. S. Gutkind, et al. (2008). "Binding-induced folding of a natively unstructured transcription factor." PLoS Comput Biol **4**(4): e1000060.

Uversky, V. N. (2002). "Natively unfolded proteins: a point where biology waits for physics." Protein Sci **11**(4): 739-56.

Uversky, V. N. (2009). "Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding." The protein journal **28**(7): 305-325.

Uversky, V. N., J. Li, et al. (2001). "Evidence for a partially folded intermediate in alpha-synuclein fibril formation." J Biol Chem **276**(14): 10737-44.

Uversky, V. N., J. Li, et al. (2001). "Trimethylamine-N-oxide-induced folding of alpha-synuclein." FEBS Lett **509**(1): 31-5.

Uversky, V. N., C. J. Oldfield, et al. (2005). "Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling." J Mol Recognit **18**(5): 343-84.

Whittington, S. J., B. W. Chellgren, et al. (2005). "Urea promotes polyproline II helix formation: implications for protein denatured states." Biochemistry **44**(16): 6269-75.

Wright, P. E. and H. J. Dyson (2009). "Linking folding and binding." Curr Opin Struct Biol **19**(1): 31-8.

Wu, Y., X. Tian, et al. (2005). "Folding of small helical proteins assisted by small-angle X-ray scattering profiles." Structure **13**(11): 1587-1597.

Zagrovic, B., J. Lipfert, et al. (2005). "Unusual compactness of a polyproline type II structure." Proc Natl Acad Sci U S A **102**(33): 11698-703.

Zheng, W. and S. Doniach (2002). "Protein structure prediction constrained by solution X-ray scattering data and structural homology identification1." Journal of Molecular Biology **316**(1): 173-187.

Zheng, W. and S. Doniach (2005). "Fold recognition aided by constraints from small angle X-ray scattering data." Protein Engineering Design and Selection **18**(5): 209-209.

IV. Conclusion

Les travaux sur la structure du P_{NTD} du VSV nous fournissent une description quantitative du désordre conformationnel qui caractérise cette région de la phosphoprotéine en solution. En particulier, les données de spectroscopie RMN indiquent la présence de structures hélicoïdales transitoires dans des proportions significatives dans les régions 1 à 12 et 25 à 31 de P_{NTD}. Les ensembles de structures sélectionnés sur base des profiles SAXS suggèrent la coexistence de deux populations de conformères compactes et étendus, qui sont affectées par la présence d'agents kosmotropiques ou chaotropiques.

Par ailleurs, un spectre HSQC ^{15}N ^{1}H réalisé sur la phosphoprotéine entière du VSV montre que les signaux correspondant au N-terminal sont identiques à ceux observés pour le peptide isolé (Figure 23), indiquant qu'aucune interaction tertiaire spécifique ne stabilise la région N-terminale dans le contexte de la phosphoprotéine entière, et suggérant donc que le P_{NTD} est également désordonné dans le contexte de la phosphoprotéine entière.

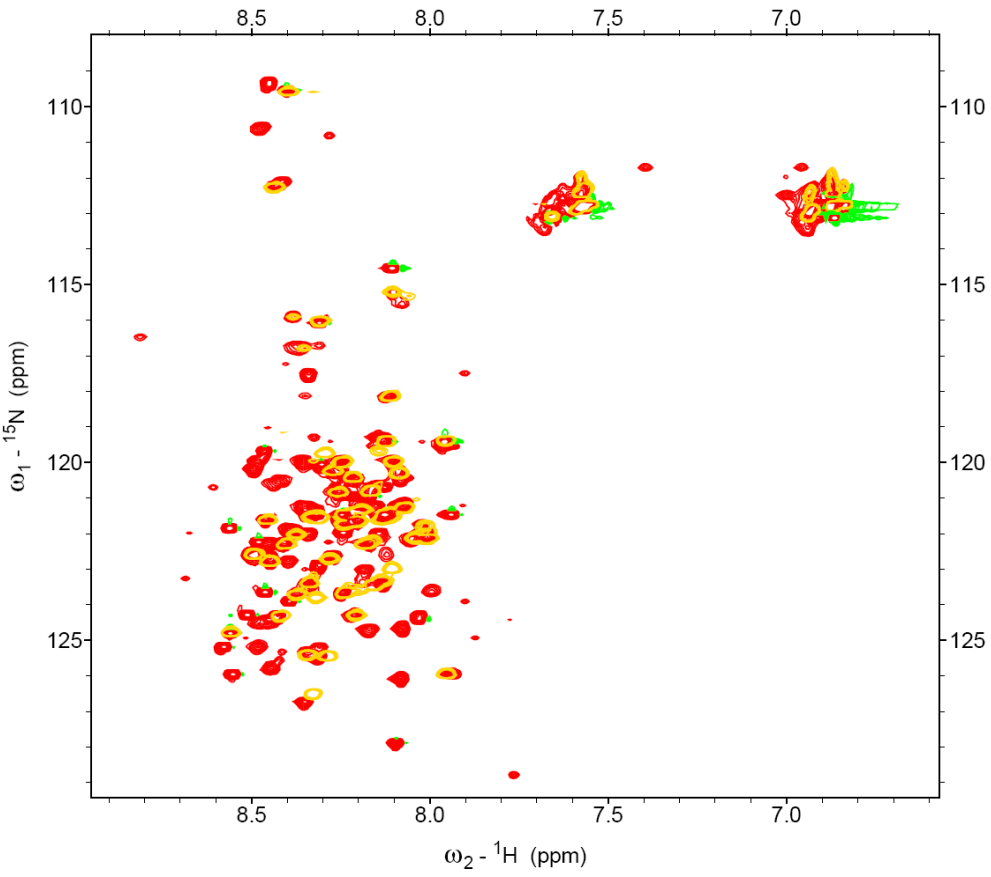


Figure 23: spectre RMN 2D [^{1}H-^{15}N]-HSQC de VSV P (rouge) superposé aux spectres [^{1}H-^{15}N]-HSQC de VSV P_{NTD} (jaune) et P_{CTD} (vert). Le spectre a été enregistré dans 20 mM Tris pH 7.5, 150 mM NaCl à 10 °C.

Concernant la région N-terminale de la phosphoprotéine du virus de la rage, les données présentées indiquent également un désordre conformationnel important, en accord avec les

résultats publiés précédemment (Gérard, Ribeiro et al. 2009). L'analyse par diffusion aux petits angles révèle une tendance marquée à la compaction et à la formation de structures secondaires à haute concentration, un phénomène qui n'est pas ou peu observé chez le VSV, et qui est probablement lié au caractère plus chargé du peptide rabique. Par ailleurs, l'étude de l'effet de la température sur la structuration du peptide, particulièrement en termes de contenu en hélices α et en polyproline II, souligne la complexité des changements structuraux induits par la température sur les protéines désordonnées, un phénomène ayant fait et faisant actuellement l'objet de nombreuses études biophysiques (Uversky, Li et al. 2001; Dawson, Muller et al. 2003; Gast, Zirwer et al. 2003; Jarvet, Damberg et al. 2003; Sanchez-Puig, Veprintsev et al. 2005; Malm, Jonsson et al. 2007; Jeganathan, von Bergen et al. 2008; Uversky 2009; Kjaergaard, Norholm et al. 2010). En particulier, outre la compaction graduelle induite par la température et qui a déjà été documentée (Nettels, Muller-Spath et al. 2009), le changement observé semble pouvoir être divisé en deux phases : un chauffage modéré (jusqu'à 40°C environ) semble induire une perte de la structure hélicoïdale résiduelle, en accord avec les observations effectuées par (Kjaergaard, Norholm et al. 2010), alors qu'à 60°C une augmentation significative de la structure hélicoïdale couplée à une perte concomitante de polyproline II semble se produire. L'occurrence simultanée de ces deux phénomènes pourrait expliquer l'apparent désaccord entre les différentes interprétations avancée dans la littérature.

INTERLUDE : MODELISATION AB INITIO DU DOMAINE CENTRAL DE LA PHOSPHOPROTEINE DU VIRUS DE LA RAGE ET COMPARAISON AVEC LA STRUCTURE CRISTALLOGRAPHIQUE

gcq#237: "If I Wanted You to Understand This, I Would Explain it Better"

(J. Cruijff, retired Dutch footballer)

Chronique d'une modélisation manquée - Modélisation de la structure du domaine central du virus de la rage et analyse à posteriori des raisons de l'échec.

La région 93-131 de la phosphoprotéine du virus de la rage avait été identifiée par la méta-prédiction de désordre comme une région structurée et la caractérisation expérimentale du fragment correspondant avait démontré qu'il s'agissait du domaine de dimérisation (Gérard, Ribeiro et al. 2009). Nous souhaitions disposer de la structure de ce domaine afin de l'incorporer dans des reconstructions de la structure entière de la P par modélisation sous forme de corps rigides articulés à partir de données de SAXS. Nous avons donc entrepris une modélisation moléculaire. Entre-temps la structure cristallographique a été obtenue (Ivanov, Crepin et al. 2010). Nous comparons ici la structure réelle avec la structure modélisée et nous discutons les raisons qui ont conduit à modéliser de manière incorrecte l'interface de dimérisation.

Nous avons utilisé une stratégie de modélisation en deux étapes, en modélisant d'abord la structure du monomère et en reconstituant ensuite le dimère. La structure du monomère de P_{CED} a été modélisée à partir de la séquence en acides aminés en utilisant le serveur LOMETS qui réalise une modélisation par homologie en utilisant une méta-analyse par enfilage (« *threading* ») (Wu and Zhang 2007). Les méthodes d'enfilage évaluent la pertinence d'enfiler une séquence protéique dans une structure connue issue d'une librairie de repliements. LOMETS réalise un consensus entre 9 programmes d'enfilage différents (Wu and Zhang 2007). Nous avons également utilisé le programme I-TASSER qui est une méthode hybride, combinant une première sélection par enfilage qui permet d'identifier un repliement consensus et une seconde étape de modélisation *ab initio* des parties non alignées (Wu, Skolnick et al. 2007; Zhang 2008; Roy, Kucukural et al. 2010). La méthode est itérative permettant de progressivement raffiner les modèles obtenus. Le monomère cristallographique, tout comme dans le meilleur modèle généré par LOMETS, contient une hélice N-terminale (aa 90-109) liée à une hélice C-terminale (aa 114-133) par une boucle (aa 110-113). L'alignement du meilleur modèle de monomère renvoyé par le serveur LOMETS avec le monomère cristallographique à l'aide du programme « secondary structure matching » (SSM) (Krissinel and Henrick 2004) indique un r.m.s.d. (root mean square deviation) de 3.0 Å. Cette différence indique une excellente correspondance entre les deux structures (Figure 24).

Dans une seconde étape, des modèles de dimère ont été construits par une procédure d'arrimage moléculaire (« *docking* ») de structures rigides avec le programme BiGGER (Krippahl, Moura et al. 2003) suivie d'une procédure de raffinement par dynamique moléculaire (Krol, Chaleil et al. 2007; Krol, Tournier et al. 2007). Parmi les 3,000 modèles de complexe sélectionnés et renvoyés par BiGGER, les orientations parallèle ou antiparallèle des monomères étaient représentées et donnaient des scores de qualité comparables. Le modèle présentant le meilleur score dans chacune des orientations a alors été soumis à une simulation de dynamique moléculaire en solvant explicite avec le logiciel GROMACS (van der Spoel, Lindahl et al. 2005) (2 x 3 = 6 simulations). Une simulation de 100 ns a été réalisée en

triplicat pour chaque modèle. La structure obtenue après 100 ns de chacun des monomères (6 x 2 = 12 monomères) a ensuite été utilisée dans une nouvelle étape d'arrimage moléculaire rigide croisé (12 x 12 = 144 simulations) générant ainsi 432,000 solutions (144 x 3000). Pour chacune des simulations, les modèles ont été reclassés en utilisant un filtre de sélection basés sur des données de mutagenèse (Gérard, Ribeiro et al. 2009) qui avaient permis d'identifier les résidus hydrophobes nécessaires pour l'auto-assemblage. Le meilleur modèle issu de chacune des 144 simulations a été minimisé en énergie avec le programme GROMACS et les modèles minimisés ont été classés en combinant trois potentiels pseudo-empiriques: Vorserv (Bernauer, Aze et al. 2007), FASTCONTACT (Champ and Camacho 2007) et EMPIRE (Liang, Liu et al. 2007). Parmi les 10 meilleurs modèles générés au bout de cette procédure, 7 présentaient une orientation parallèle des monomères et 3 présentaient une orientation antiparallèle. La comparaison de ces modèles de dimères avec la structure cristallographique a indiqué des valeurs de r.m.s.d. élevées indiquant la mauvaise qualité de la modélisation. Dans la structure cristallographique, les monomères sont inclinés l'un par rapport à l'autre selon un angle de 46° proche de l'angle de 50° généralement rencontrés lors de l'assemblage hélice-hélice dans des protéines globulaires mais différent de l'angle de 20°C généralement observés dans les faisceaux de 4 hélices (*« 4 helix bundles »*). Les structures de dimères obtenues par modélisation présentaient une orientation incorrecte entre les monomères. Dans ce cas, l'échec de la modélisation vient de la sélection arbitraire de seulement deux modèles après la première étape d'arrimage rigide dictée par le temps de calcul requis pour réaliser la procédure de raffinement. Les modèles choisis présentaient déjà une orientation incorrecte des monomères qui n'a pas pu être corrigée par la procédure de raffinement (Figure 24). En effet, le temps simulé dans la dynamique moléculaire était trop court pour permettre aux deux monomères de se séparer et de se réassocier dans une autre orientation (Figure 24). Néanmoins, on constate que la plupart des simulations montrent un dépliement partiel des hélices α qui suggèrent que les structures modélisées sont instables. Ce problème aurait pu être évité si nous avions réalisé une première analyse de l'orientation des monomères au bout de la première étape d'arrimage.

Il est également intéressant de noter que la soumission au serveur I-TASSER de deux copies en tandem de la séquence de P_{CED} séparées par un cours segment flexible (« *linker* ») constitué d'une répétition de Ser et de Gly renvoie des modèles composés d'un arrangement compact d'hélices α formant une large interface hydrophobe (données non-présentées). Cette modélisation confirme la tendance d'auto-assemblage du monomère.

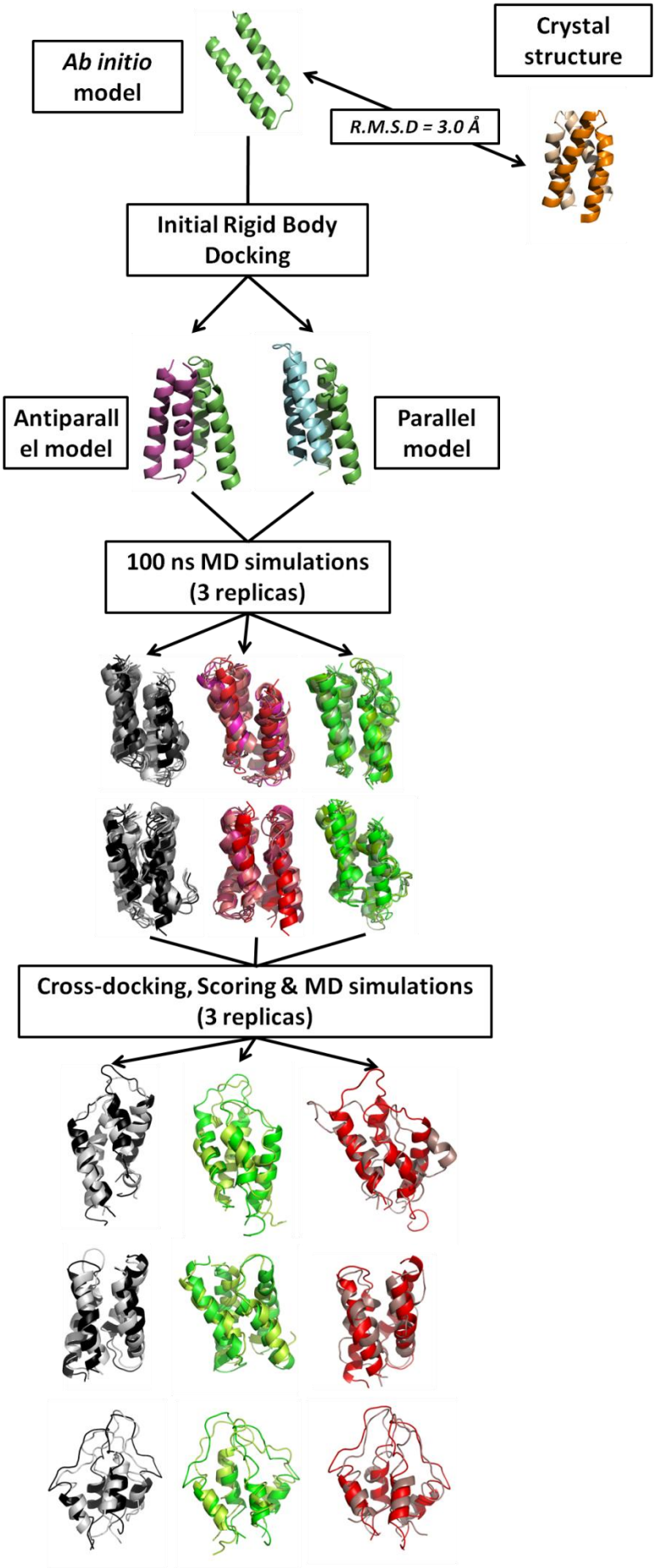


Figure 24 : Modélisation *ab initio* du domaine de dimérisation de la phosphoprotéine du Virus de la Rage (RV P_{CED}) : La procédure repose sur l'utilisation d'un modèle initial du monomère correspondant aux résidus 90 à 133 de la phosphoprotéine obtenu *via* le serveur LOMETS (Wu and Zhang 2007). L'alignement avec un monomère extrait de la structure cristallographique du domaine (code pdb : 3L32, (Ivanov, Crepin et al. 2010)) réalisée à l'aide du serveur SSM (Krissinel and Henrick 2004) indique un R.M.S.D de 3.02 Å. Après une première étape de docking rigide (BiGGER, (Krippahl, Moura et al. 2003)), les deux modèles dimériques affichant le meilleur score global dans les orientations parallèles et antiparallèles sont utilisés comme conformations initiales dans des simulations de Dynamique Moléculaire en solvant explicite de 100 ns réalisées en triplicats dans GROMACS ((van der Spoel, Lindahl et al. 2005). 2 x 3 = 6 simulations sont réalisées et les conformations de chaque monomère après 100 ns de simulation (6 x 2 monomères) sont extraites et utilisées dans un nouveau cycle de docking rigide (12 x 12 = 144 runs) générant ainsi 144 x 3000 = 432000 solutions. Après reclassement des modèles sur base de données de mutagénèse dirigée concernant l'interface d'interaction (Gérard, Ribeiro et al. 2009), le dimère obtenant le meilleur score global avec chacune des combinaisons de conformères de P_{CED} est extrait et minimisé en énergie dans GROMACS. Les 144 modèles dimériques ainsi raffinés sont reclassés en combinant trois potentiels pseudo-empiriques différents (Vorserv (Bernauer, Aze et al. 2007), FASTCONTACT (Champ and Camacho 2007) et EMPIRE (Liang, Liu et al. 2007)). Les 10 meilleurs modèles obtenus consistent en 7 modèles parallèles et 3 modèles antiparallèles, indiquant que l'agencement parallèle au sein du dimère est nettement favorisé, comme cela est observé dans la structure cristallographique. Enfin, les 3 meilleurs modèles sont à nouveaux simulés en triplicats pendant 100 ns.

CHAPITRE IV : STRUCTURE DU DOMAINE C-TERMINAL DE LA PHOSPHOPROTEINE DU VIRUS DE LA STOMATITE VESICULAIRE

gcq#256: "Fly to the Court of England and Unfold"

(Macbeth, Act 3, Scene 6, William Shakespeare)

I. Introduction

Le domaine C-terminal de la phosphoprotéine (P_{CTD}) constitue un des modules structuraux et fonctionnels les plus conservés de la phosphoprotéine des *Rhabdoviridae* (Gérard, Ribeiro et al. 2009), et est notamment impliqué dans la fixation aux complexes nucléoprotéine-ARN. Les travaux présentés dans l'article ci-après concernent la structure du domaine C-terminal de la phosphoprotéine du Virus de la Stomatite Vésiculaire, résolue par RMN grâce aux efforts collaboratifs conjoints des docteurs Euripedes de Almeida Ribeiro Jr, en post-doctorat dans l'équipe, et Adrien Favier du laboratoire de Résonance Magnétique Nucléaire de l'Institut de Biologie Structurale de Grenoble. La structure RMN révèle un agencement des éléments de structure secondaire au sein du domaine P_{CTD} similaire à celui observé chez RV, avec un « fold » néanmoins plus épuré qui confère au domaine une plus grande flexibilité qui explique l'échec des différents essais de cristallisation (Euripedes de Almeida Ribeiro Jr, communication personnelle).

Ma contribution à ce travail consista principalement à traduire la conservation des résidus du P_{CTD} sur la structure tridimensionnelle (Pei and Grishin 2001; Meng, Pettersen et al. 2006), et à comparer la topologie du domaine de VSV avec la structure cristallographique disponible pour le P_{CTD} du RV (Krissinel and Henrick 2004).

Plusieurs fonctions ont été attribuées au domaine P_{CTD} du VSV en utilisant des délétions C-terminales de la phosphoprotéine, et notamment la fixation à N-ARN et L (Paul, Chattopadhyay et al. 1988; Takacs, Das et al. 1993). Des simulations de dynamique moléculaire réalisées sur le P_{CTD} du VSV et sur deux constructions délétées des 11 (ΔC11) ou des 21 (ΔC21) derniers résidus sont présentées en complément à l'article. Ces données fournissent une analyse théorique de la stabilité du domaine, et suggèrent que la délétion ΔC21 entraîne une déstabilisation importante du P_{CTD} incitant à interpréter avec prudence les données biochimiques utilisant des délétions C-terminales de la phosphoprotéine chez le VSV (Paul, Chattopadhyay et al. 1988; Takacs, Das et al. 1993).

II. ARTICLE IV: Solution Structure of the C-Terminal Nucleoprotein–RNA Binding Domain of the Vesicular Stomatitis Virus Phosphoprotein

Auteurs: Euripedes de Almeida Ribeiro Jr, Adrien Favier, Francine Gérard, Cédric Leyrat, Bernhard Brutscher, Danielle Blondel, Rob Ruigrok, Martin Blackledge et Marc Jamin.

Publié en 2008 dans *Journal of Molecular Biology*, Volume 382, pages 525-538

doi:10.1016/j.jmb.2008.07.028 J. Mol. Biol. (2008) 382, 525–538

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Solution Structure of the C-Terminal Nucleoprotein–RNA Binding Domain of the Vesicular Stomatitis Virus Phosphoprotein

Euripedes A. Ribeiro Jr[1]†, Adrien Favier[2]†, Francine C. A. Gerard[1], Cédric Leyrat[1], Bernhard Brutscher[2], Danielle Blondel[3], Rob W. H. Ruigrok[1], Martin Blackledge[2]* and Marc Jamin[1]*

[1]*UJF-EMBL-CNRS—UMR 5233—Unit of Virus Host Cell Interactions, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France*

[2]*Institut de Biologie Structurale Jean-Pierre Ebel, CNRS, CEA, UJF, UMR 5075, 41 rue Jules Horowitz, F-38027 Grenoble, Cedex, France*

[3]*Laboratoire de Virologie Moléculaire et Structurale, UMR 2472 CNRS-1157 INRA Bât. 14B, 1 Avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France*

Received 10 June 2008; accepted 7 July 2008
Available online
16 July 2008

Beyond common features in their genome organization and replication mechanisms, the evolutionary relationships among viruses of the *Rhabdoviridae* family are difficult to decipher because of the great variability in the amino acid sequence of their proteins. The phosphoprotein (P) of vesicular stomatitis virus (VSV) is an essential component of the RNA transcription and replication machinery; in particular, it contains binding sites for the RNA-dependent RNA polymerase and for the nucleoprotein. Here, we devised a new method for defining boundaries of structured domains from multiple disorder prediction algorithms, and we identified an autonomous folding C-terminal domain in VSV P (P_{CTD}). We show that, like the C-terminal domain of rabies virus (RV) P, VSV P_{CTD} binds to the viral nucleocapsid (nucleoprotein–RNA complex). We solved the three-dimensional structure of VSV P_{CTD} by NMR spectroscopy and found that the topology of its polypeptide chain resembles that of RV P_{CTD}. The common part of both proteins could be superimposed with a backbone RMSD from mean atomic coordinates of 2.6 Å. VSV P_{CTD} has a shorter N-terminal helix (α_1) than RV P_{CTD}; it lacks two α-helices (helices α_3 and α_6 of RV P), and the loop between strands β_1 and β_2 is longer than that in RV. Dynamical properties measured by NMR relaxation revealed the presence of fast motions (below the nanosecond timescale) in loop regions (amino acids 209–214) and slower conformational exchange in the N- and C-terminal helices. Characterization of a longer construct indicated that P_{CTD} is preceded by a flexible linker. The results presented here support a modular organization of VSV P, with independent folded domains separated by flexible linkers, which is conserved among different genera of *Rhabdoviridae* and is similar to that proposed for the P proteins of the *Paramyxoviridae*.

Keywords: phosphoprotein; vesicular stomatitis virus; rabies virus; NMR structure; virus replication

Edited by M. F. Summers

**Corresponding authors.* E-mail addresses: martin.blackledge@ibs.fr; jamin@embl.fr.

† E.A.R. and A.F. contributed equally to this work.

Abbreviations used: P, phosphoprotein; VSV, vesicular stomatitis virus; RV, rabies virus; *MNV*, *Mononegavirales*; N, nucleoprotein; L, RNA polymerase; NC, nucleocapsid; IND serotype, Indiana serotype; NJ serotype, New Jersey serotype; SEC, size-exclusion chromatography; MALLS, multiangle laser light scattering; RI, refractometry; HSQC, heteronuclear single quantum coherence; PDB, Protein Data Bank; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; RDC, residual dipolar coupling.

Introduction

Vesicular stomatitis virus (VSV) and rabies virus (RV) belong to the *Rhabdoviridae* [order: *Mononegavirales* (*MNV*)] and have a nonsegmented negative-strand RNA molecule as genome. VSV causes infections in cattle and horses, rarely in humans, but it bears a high economical impact on the farm industry, in particular in the Americas where the virus is endemic.[1] For many years, VSV has served as a model system for studying the mechanism of replication of negative-strand RNA viruses. Indeed, *MNV* viruses vary widely in morphology and

0022-2836/$ - see front matter

cellular interactions, but they share a similar organization of their genomes as well as similar modes of RNA replication and transcription.[2,3] The viral genome (9 to 18 kb) comprises five common genes [successively coding for the nucleoprotein (N), the phosphoprotein (P), the matrix protein, the glycoprotein, and the RNA polymerase (L)] that fulfill similar roles in the life cycles of these viruses. Both VSV and RV use only three of these viral proteins, N, P, and L, to synthesize viral RNA in an efficient and regulated manner.[4,5] The genomic RNA is always encapsidated by N, forming a helical nucleocapsid (NC) that serves as a template for replication and transcription.[6] P is an essential component in the replication machineries of all *Rhabdoviridae*, and it plays multiple roles during the viral replication cycle. In the early stages of viral infection, P forms a complex with N, named the N^0–P complex, which maintains N in a soluble and monomeric form, prevents the binding of N to cellular RNA, and, therefore, preserves N for encapsidating the viral RNA. In addition, P acts as a cofactor of the viral RNA-dependent L in both transcription and replication processes.[5,7] L carries out all the enzymatic activities but is unable to bind to NCs. P binds to both NCs and L and allows the association of the polymerase with its template.

Two stable domains of VSV P have been identified by limited proteolysis.[8,9] The isolated central domain (from amino acid 107 to 177) forms a dimer whose structure was solved by X-ray crystallography.[10] Recently, we have shown that full-length P from both VSV and RV also forms dimers in solution.[11] Another stable domain was found at the C-terminal extremity of P (P_{CTD}).[9] Mutation and deletion studies with VSV P from both the Indiana (IND) and New Jersey (NJ) serotypes identified several functional regions within P_{CTD}. At the N-terminal extremity of P_{CTD}, one region (amino acids 201–210) is critical for RNA replication and another one (amino acids 191–200) is involved in virus assembly, although the exact roles of these regions are unknown.[12] The C-terminal part of P_{CTD} is necessary for binding to N and in particular to N–RNA templates,[9,13–16] and the central region of P_{CTD} is necessary for binding to L and contains phosphorylation sites.[13,15–19] No structural information about this domain of VSV P is available, but the three-dimensional (3D) structure of the C-terminal domain of RV P (RV P_{CTD}) was solved by X-ray crystallography.[20] Essential residues for transcription and replication were identified by mutational studies in the homologous protein from Mokola virus and mapped onto two regions of RV P_{CTD}, one on each face of the molecule. VSV and RV belong to two different genera within the *Rhabdoviridae* family, the *Vesiculovirus* and *Lyssavirus* genera, respectively. Multiple alignments of P from these genera, however, reveal no homology (0% identity, 4% similarity). In the *Paramyxoviridae*, another family of *MNV*, the C-terminal domain of P also binds to N–RNA complexes. The 3D structures of the C-terminal domain of P from viruses belonging to two different genera of *Paramyxoviridae*, measles virus and Sendai virus, share a similar structure,[21,22] which is different from that of the C-terminal domain of RV P. This observation raises the question of structural conservation of P proteins within the *Rhabdoviridae* family.

In this work, we analyzed the amino acid sequence of VSV P (IND) to localize the boundary of P_{CTD}, and we showed that the identified C-terminal domain of P constitutes an autonomous folding unit. We determined its 3D structure and analyzed its dynamic properties by NMR spectroscopy. The structure revealed homology with the C-terminal domain of RV and is discussed in the light of previous functional data. The results presented here support a common modular organization of P proteins within the *Rhabdoviridae* family.

Results

The modular organization of VSV P

P proteins from the *Rhabdoviridae* are intrinsically disordered proteins that consist of structured domains alternating with disordered regions.[23] The location of disordered regions in the sequence of VSV P (IND) was predicted from the amino acid sequence by using 16 different algorithms available through web servers. The predictions showed large differences in the number, position, and length of the predicted disordered regions (data not shown). In order to define consensus locations for the boundaries between structured and disordered regions, we devised a scoring method based on a simple binary voting procedure that integrated the results from all disorder predictions. The calculated score (Fig. 1a) varied from 0 to 1, and, for the sake of simplicity, we arbitrarily defined two threshold levels. Residues with a score equal to or below 0.50 were considered disordered, and residues with a score equal to or above 0.75 were considered structured. Residues with a score in the intermediate zone, between 0.50 and 0.75, were considered ambiguous. For VSV P, the disorder score revealed the presence of three disordered regions (amino acids 39–84, 109–131, and 187–195) and four structured regions (amino acids 1–19, 95–105, 137–177, and 216–265) (Fig. 1b). A similar pattern of disordered and structured regions was found with the amino acid sequence of VSV NJ serotype, which is longer (274 aa) than that of the IND serotype (265 aa) and exhibits two small gaps at positions 45 and 71–73 as well as a large insertion of 14 residues between residues 195 and 196 (numbering according to the IND serotype). The main insertion in the C-terminal half of the NJ serotype was predicted to be disordered, resulting in a longer C-terminal disordered region than in the IND serotype (data not shown).

On the basis of this prediction, three different constructs were generated to locate the boundaries of a stable and structured VSV P_{CTD} domain (Fig. 1c). The shortest construct (P_{CTD1}: amino acids 215–265)

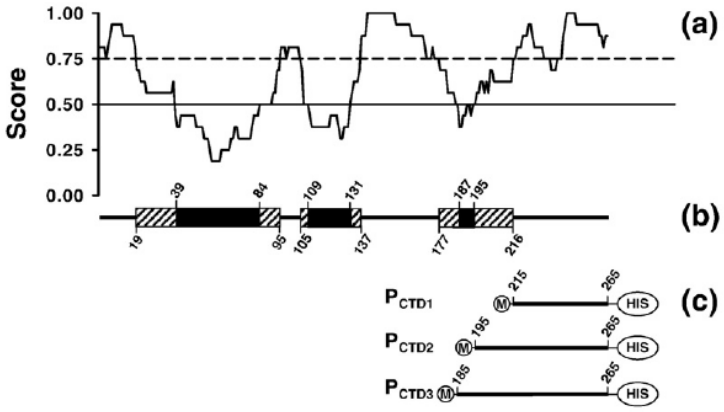


Fig. 1. (a) Modular organization of VSV P. The graph shows the normalized score for disorder as a function of residue number calculated from the results of 16 different predictors. The score was calculated by assigning, for each prediction, a value of 0 to the residues predicted to be disordered and a value of 1 to those predicted as structured, by adding these values for each residue, and by normalizing the results. Residues with a normalized score ≤0.5 are defined as consensus disordered regions; residues with a normalized score ≥0.75 are defined as consensus structured regions. (b) Consensus location of disordered regions. The scheme shows the location of consensus disordered regions with a score below 0.50 (black boxes) and with a score ranging between 0.75 and 0.50 (striped boxes). Numbers above the boxes indicate the limits of the consensus disordered regions and numbers below indicate the limits of structured regions. (c) Schematic representation of the constructs of the VSV C-terminal domain of P used in this article. Circled M represents the added N-terminal methionine, and circled HIS represents the His-tag sequence (LEHHHHHH).

was mainly found in inclusion bodies, and after purification on a Ni^{2+} column in an unfolded form (6 M urea), the protein could not be properly refolded. The intermediate construct (P_{CTD2}: amino acids 195–265) was mainly expressed in soluble form, whereas the longest construct (P_{CTD3}: amino acids 185–265) was found in inclusion bodies although a significant fraction was soluble. Therefore, the structural work was carried out mainly with P_{CTD2}. It should also be noted that P_{CTD3} obtained by refolding from inclusion bodies was undistinguishable from that obtained in the soluble fraction.

The C-terminal domain of VSV P (P_{CTD}) is an autonomous folding unit

The molecular masses measured by mass spectrometry for P_{CTD2} (9070±2 Da) and its ^{15}N-labeled form (9188±2 Da) indicated that it lacked the N-terminal methionine (molecular mass calculated from the gene without the N-terminal Met is 9075 and 9191 Da, respectively). These results were confirmed by N-terminal sequencing. Conversely, P_{CTD3} was obtained with an N-terminal methionine, in agreement with the observed preferences of the methionyl aminopeptidase of *Escherichia coli* for the penultimate amino acid.[24] In P_{CTD2}, the second amino acid is a serine, one of the most processed penultimate amino acids, whereas in P_{CTD3}, it is an asparagine, one of the least processed.[24]

The corresponding domain of RV P binds to RV NCs.[20] The binding of the C-terminal domain of VSV P to viral NCs was tested by centrifugation through a glycerol cushion and SDS-PAGE. NCs isolated from virus-infected cells sedimented through the glycerol and were found in the pellet (lanes 2 and 3, Fig. 2), whereas VSV P_{CTD2} remained in the supernatant (lanes 6 and 7, Fig. 2). When P_{CTD2} and NCs were previously incubated, both proteins cosedimented, and the C-terminal domain was found both in the supernatant and in the pellet (lanes 4 and 5, Fig. 2). Similar results were obtained with P_{CTD3}, demonstrating that our fragments of VSV P conserved their ability to bind to the viral NCs, as previously shown for another construct (amino acids 192–265).[9]

Analysis by size-exclusion chromatography (SEC) using online detection by multiangle laser light scattering (MALLS) and refractometry (RI) indicated that P_{CTD2} and P_{CTD3} are monodisperse in solution. Their weight-averaged molecular masses of 9700± 800 Da and 11,800±300 Da (Fig. S1a and b) showed that both proteins are monomeric (molecular mass

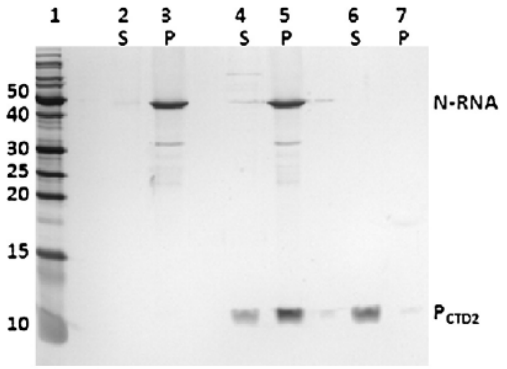


Fig. 2. Binding of the C-terminal domain of VSV P_{CTD2} to NCs. NCs (N–RNA complex) purified from infected BSR cells by CsCl gradient centrifugation were incubated with P_{CTD2} (amino acids 195–265). After centrifugation of the mixture through a 15% (v/v) glycerol cushion, samples of supernatant (S) and pellet (P) were separated by 15% SDS-PAGE and detected by silver staining. NCs were used at a final concentration of 5 μM. Lane 1, protein standards with molecular mass indicated on the left; lanes 2 and 3, N–RNA alone; lanes 4 and 5, N–RNA:P_{CTD2} in a 1:1 molar ratio; lanes 6 and 7, P_{CTD2} alone.

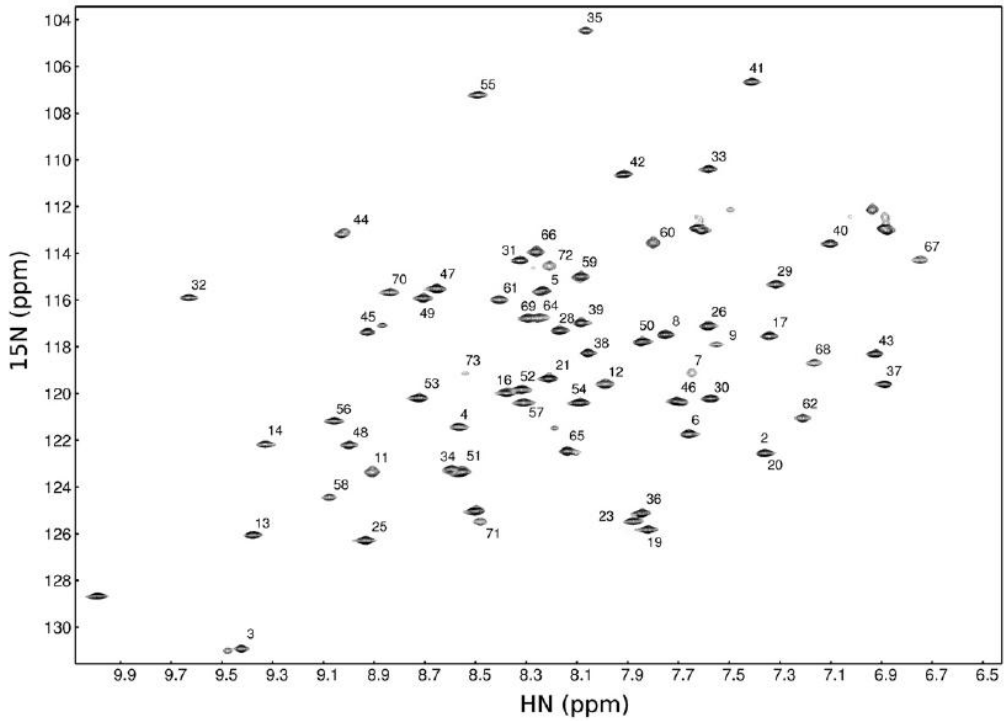


Fig. 3. The ^{1}H–^{15}N HSQC spectrum of P_{CTD2}. The spectrum was measured at 800 MHz and at 20 °C. Residues are numbered P_{CTD2} starting at amino acid 195.

calculated from the gene is 9075 and 10,258 Da, respectively) as previously suggested by SEC analysis,[9] in contrast to full-length VSV P that forms dimers.[11] Also, a value of 9700±500 Da was obtained for P_{CTD2} from small-angle neutron scattering measurements (data not shown).

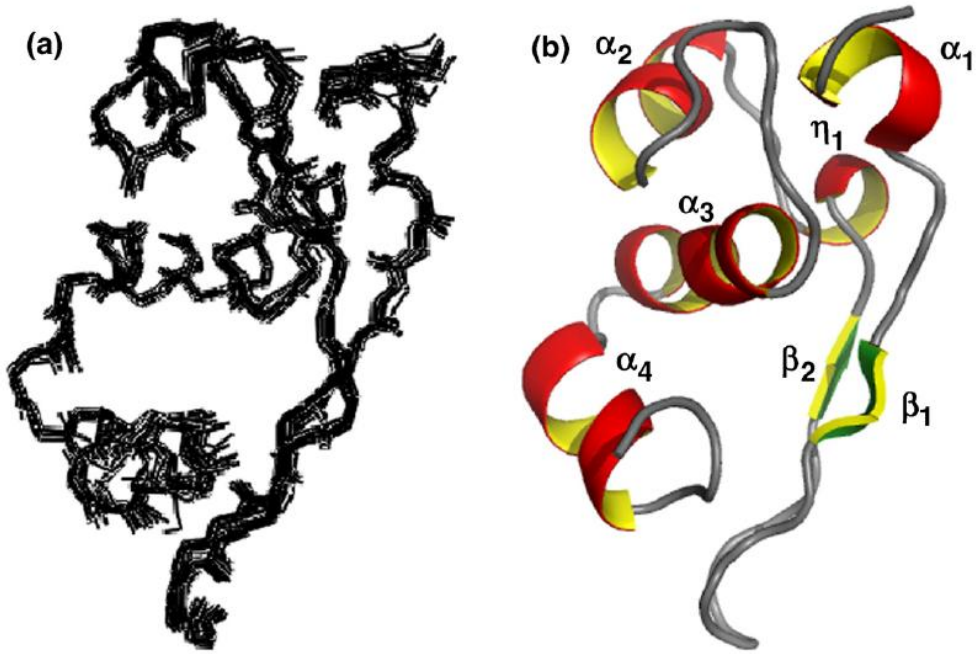


Fig. 4. NMR structure of P_{CTD2}. (a) Ensemble of the 20 best structures calculated from NMR data (PDB accession number: 2k47). (b) Cartoon representation of the lowest-energy NMR structure of VSV P_{CTD2}. Secondary-structure elements are colored and labeled. The figure was prepared with the PyMOL program.

Recombinant P_{CTD2} and P_{CTD3} were folded as shown by far-UV circular dichroism (CD) (Fig. S2a) and by 1D NMR[25] (data not shown) and exhibited a similar amount of secondary structure although P_{CTD3} contains 10 additional residues. Figure 3 shows the ^{1}H–^{15}N heteronuclear single quantum coherence (HSQC) spectrum of P_{CTD2} at pH 7.5 and 20 °C. In these conditions, both P_{CTD2} and P_{CTD3}, prepared as described above, were stable according to HSQC spectra over a period of 6 weeks, and thermal denaturation of both proteins monitored by CD at 222 nm was fully reversible (Fig. S2b). The protein denatured in a single cooperative transition with a T_m near 45 °C for P_{CTD2} and near 47 °C for P_{CTD3}. The far-UV CD spectra recorded at 25 °C after one cycle of denaturation/renaturation were superimposed to the initial spectra obtained at 25 °C (Fig. 3a).

NMR structure

VSV P_{CTD} resembles the structure of RV P_{CTD}

The 3D structural ensemble defining P_{CTD2} is shown in Fig. 4a, and the associated statistical analysis is presented in Table 1. VSV P_{CTD2} forms a single compact domain that has the shape of a triangular prism. It is composed of four α-helices (α_1–α_4), one 3^{10} helix (η_1), and two short β-strands (β_1–β_2) that form an antiparallel β-sheet (Fig. 4b). A search of the Protein Data Bank (PDB) using the programs DALI[26] or SSM (secondary-structure matching)[27] revealed no hit with a Z score higher than 2.5 and did not detect the P_{CTD} of RV.[20] However, structural alignment with RV P_{CTD} using the SSM program[27] gave a Z score of 2.3 and an RMSD value of 2.6 Å over the 44 aligned residues. Figure 5a and b show the superposition of both structures. The short α_1 helix, the two β-strands β_1 and β_2, the 3^{10} helix η_1, and the helices α_2, α_3, and α_4 of VSV P_{CTD} match with corresponding secondary-structure elements in RV P_{CTD} (α_1, β_1, β_2, η_1, α_2, α_4, and α_5, respectively). Comparison of the topology of the peptide chain between VSV and RV P_{CTD} indicated a similar organization in both proteins (Fig. 5c and d). Helices α_3 and α_6 of RV P are missing in VSV P, helices α_1 and α_2 are shorter in VSV than in RV P, while the loop connecting β_1 and β_2 is longer in VSV than in RV P (Fig. 5e).

NMR spectrum of P_{CTD3}

The HSQC spectrum obtained for P_{CTD3} was similar to that obtained for P_{CTD2} (Fig. S3a). Only three additional resonances were visible in the amide backbone region, with ^{1}H chemical shifts in the range associated with disordered structure. The seven other additional residues did not give rise to observable amide backbone resonances, indicating the presence of intermediate conformational exchange, while N-terminal sequencing performed on the sample used for the NMR experiments confirmed the integrity of the protein from its N-terminal methionine residue. Some resonances in the spectrum of P_{CTD2} corresponding to residues 197 and 198 in helix α_1 and to residues 237 to 240 in the loop connecting helices α_2 and α_3 were slightly shifted in P_{CTD3} spectrum. These residues are the closest in space to the N-terminal extension in P_{CTD3} and the most susceptible to be affected by the presence of an additional flexible N-terminal extension (Fig. S3b).

Table 1. NMR structure parameters, data statistics,[a] and details of structure determination

Number of experimental distance constraints	
All	1091
Intraresidual	227
Sequential ($\|i-j\|=1$)	313
Short range ($1<\|i-j\|\leq 5$)	231
Long range ($\|i-j\|>5$)	320
RMSDs from experimental distance constraints (Å)	
NOEs	0.14±0.03
Number of NOE violations	
Violations >0.5 Å	14±4
Violations >0.3 Å	32±8
Violations >0.1 Å	75±17
RMSDs from experimental dihedral constraints (°)[b]	
ϕ/ψ	(68) 0.19±0.05
Reduced chi-square (χ^2/N) from experimental RDC restraints[b]	
C^{α}–H^{α}	(30) 2.1±0.2
N–H^{N}	(40) 1.0±0.1
C^{α}–C′	(36) 0.7± 0.1
Alignment tensor eigenvalues	
Da	$13.0\pm0.2\times10^{-4}$
Dr	$3.3\pm0.1\times10^{-4}$
Deviation from the idealized covalent geometry	
Bonds (Å)	0.0039±0.0001
Angles (°)	0.48±0.006
Impropers (°)	0.42±0.01
Energies (kcal/mol)	
Total	−2430±85
van der Waals	−691±8
Electrostatic	−2217±82
Coordinate precision (Å)	
Backbone secondary structure[c]	0.27±0.06
Heavy atoms secondary structure	0.88±0.09
Backbone	0.43±0.09
Heavy atoms	0.93±0.11
Percentage of non-Gly and non-Pro residues in Ramachandran regions[d]	
Core	77.7
Allowed	21.2
Generous	1.2
Disallowed	0.0

[a] The statistics were obtained from a subset of the 20 best structures following the protocol described in Materials and Methods. The number of the various constraints is given in parentheses.
[b] The dihedral angle constraints are inferred from the program TALOS (reference).
[c] Secondary structural elements defined as (10–13), (23–26), (28–30), (34–41), (49–59), and (65–68).
[d] These values are calculated with the program PROCHECK-NMR.

Conformational dynamics

NMR spin relaxation is sensitive to motions of the associated nuclei and can be used to probe local and

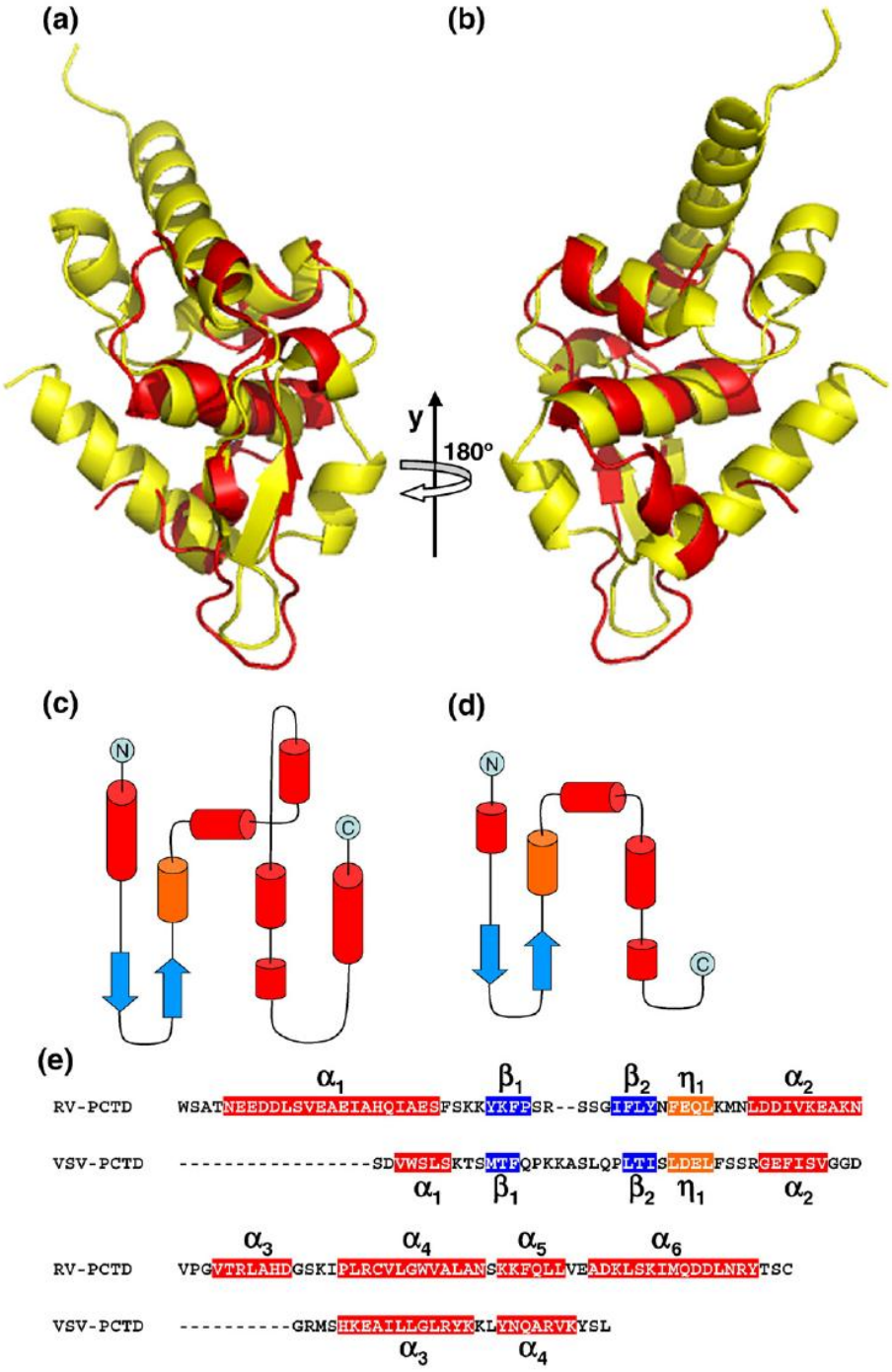


Fig. 5. Comparison of VSV P_{CTD} and RV P_{CTD}. (a and b) Structural alignments. In yellow, ribbon representation of RV P_{CTD} (PDB accession number: 1VYI); in red, ribbon representation of VSV P_{CTD2}. The RMSD is 2.6 Å for the 44 superposed residues. (b) shows the structure rotated by 180° around the *y*-axis. (c and d) Topology of RV (c) and VSV (d) P_{CTD}. (e) Alignment of the amino acid sequences of RV and VSV PCTD based on structural alignment.

global protein dynamics. The relaxation rates of longitudinal and transverse magnetization of backbone ^{15}N nuclei (R1 and R2), as well as ^{1}H–^{15}N steady-state heteronuclear nuclear Overhauser enhancement (NOE), are sensitive to motions on a picosecond to nanosecond timescale. We present experimental relaxation rates rather than a Lipari–Szabo analysis[28] because the apparently significant exchange contributions in the terminal helices, combined with restricted angular sampling in the secondary-structure elements of the molecule, preclude an accurate analysis of the rotational diffusion

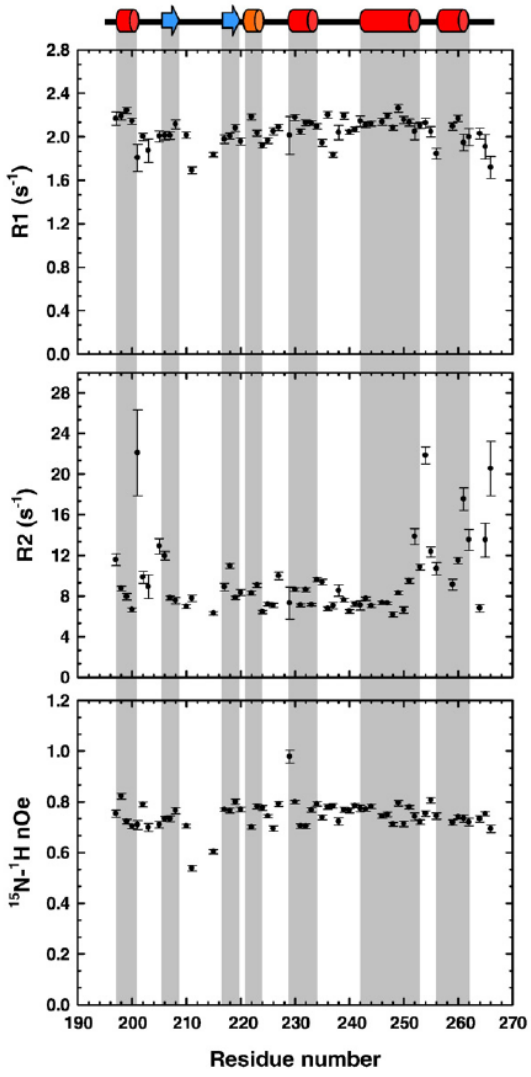


Fig. 6. Dynamics of P_{CTD2}. Relaxation rates of longitudinal, R1 (a), and transverse, R2 (b), magnetization of backbone ^{15}N nuclei and ^{1}H–^{15}N heteronuclear NOE (c). Secondary-structure elements are represented above the plots and by gray shades within the plots.

tensor. The R1 and R2 relaxation rates and heteronuclear NOEs are shown in Fig. 6 for each backbone amide in VSV P_{CTD2}. In addition to enhanced flexibility in the region of residues 209 to 215 as indicated from the heteronuclear NOE experiment, transverse relaxation rates were most probably affected by conformational exchange on the microsecond to millisecond timescale in the N-terminal (amino acids 195–206) and in particular in the C-terminal (amino acids 251–265) helical region. The protein used for NMR spectroscopy contained the His-tag, but no resonances were observed for this part of the protein, also suggesting a conformational exchange on the microsecond to millisecond timescale. In a similar study of the dynamic behavior of VSV P_{CTD3}, essentially identical heteronuclear NOE values were measured throughout the protein for the common regions, indicating that the additional 10 amino acids have a negligible effect on the stability of the protein on the observed timescales. For the three additional amide backbone resonances observed in the NMR spectrum of P_{CTD3}, one had a heteronuclear NOE value near zero and the other two exhibited negative values, indicating an important flexibility in this region of the protein.

Discussion

Modular organization of P

Two observations reported here support a modular organization of VSV P. Firstly, in a modular protein, independent domains are expected to fold in an autonomous process, and we found that recombinant VSV P_{CTD} folds in *E. coli* and that its unfolding is fully reversible. The shortest construct (P_{CTD1}), lacking helix α_1 and strand β_1, could not be expressed as a soluble protein and could not be refolded. Also, the central dimerization domain folds independently of the rest of the protein.[10] Secondly, these two autonomous folding units seem to be attached by a flexible linker. The short region preceding helix α_1 in P_{CTD3} was predicted to be disordered by our scoring method (Fig. 1a), and NMR spectroscopy confirmed that this region of P is flexible. Three additional amide backbone resonances were visible in the 2D HSQC NMR spectrum obtained with P_{CTD3} as compared to the spectrum obtained with P_{CTD2}. Heteronuclear NOEs for these three residues had null (one residue) or negative values (two residues), indicating a high degree of flexibility. The absence of peaks for the other residues probably resulted from conformational exchange on a timescale similar to that of the NMR experiment (microseconds to milliseconds). In P proteins from members of the *Paramyxoviridae* family, the C-terminal N–RNA binding domain pX is also separated from the central tetramerization domain by a flexible linker,[21,29] indicating that the mobility of P_{CTD} is conserved between different *MNV* families and suggesting that it is required for its biological function. Also, it is noteworthy to mention that a longer disordered region, separating the central dimerization domain from the N–RNA binding C-terminal domain, is predicted in VSV P from the NJ serotype.

Mapping of the functional regions within VSV P_{CTD}

VSV P is a multifunctional protein and its C-terminal domain (P_{CTD}) participates in different processes during viral replication. Mutation and deletion studies have been used to identify amino acids or regions that are involved in the different activities of P_{CTD} (Fig. 7a). This domain binds to N in the NC and to L[14,16,17,30] but was also shown to participate in virus assembly, possibly by interacting with other viral or cellular proteins.[12] The last 21 C-terminal residues of P (amino acids 235–265) are

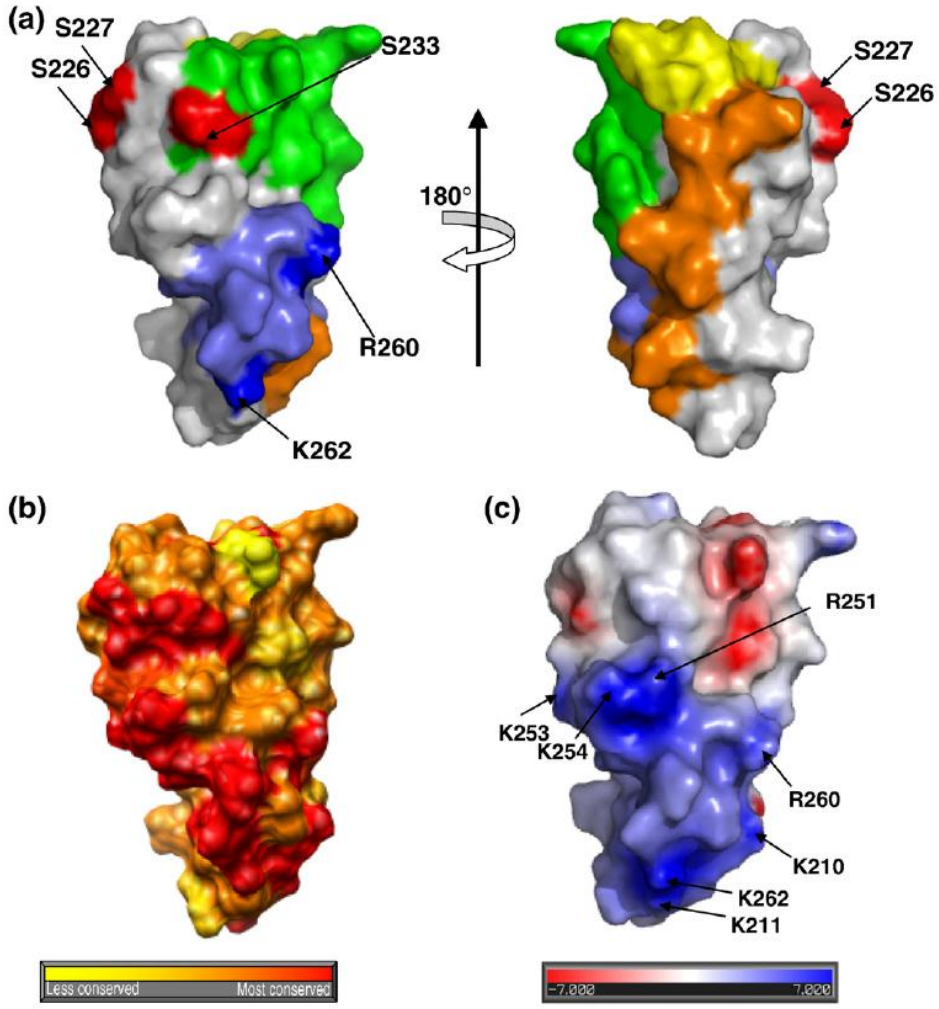


Fig. 7. Mapping of functional residues, electrostatic surface potential, and amino acid conservation on the surface of VSV P_{CTD2}. (a) Functional mapping. Surface representations of P_{CTD2} showing the residues involved in virus assembly in yellow (amino acids 195–200), those involved in RNA replication in orange (amino acids 201–210), those involved in binding to L in green (amino acids 231–244), those involved in binding to N in blue (amino acids 255–265 with Arg260 and Lys262 in dark blue), and the phosphorylation sites in red (Ser226, Ser227, and Ser233). (b) Amino acid sequence conservation among VSV strains. The conservation score calculated with the AL2CO program is mapped on the surface of P_{CTD2} using the Chimera program, with red denoting the most conserved residues and yellow indicating the least conserved residues. (c) Electrostatic surface potential. The surface potential was calculated with the Delphi program and is color-coded on the surface from red (negatively charged residues; $-12\ k_BT$) to blue (positively charged residues; $+12\ k_BT$).

required for transcriptional activity and for the binding to NC templates.[16,17,30] This interacting region was narrowed to the last 10 or 11 C-terminal residues.[16,30] Deletion of the last 11 C-terminal residues of VSV P almost completely abrogated viral transcription and replication and significantly reduced the binding to N–RNA complexes, to soluble N, and to L[16] (amino acids 255 to 265 are shown in blue in Fig. 7a). Two critical basic residues, Arg260 and Lys262, were identified in this C-terminal region. Mutation of Arg260, a residue strictly conserved among all *Vesiculoviruses*, is sufficient for significantly reducing the binding of P to N–RNA, while mutation of Lys262 has almost no effect. However, mutations of Arg260 or of Lys262 have a strong effect on *in vitro* transcription and on binding to L but has no effect on RNA replication or on binding to soluble N.[16] These results underline the importance of C-terminal residues in binding NCs and L. Another region, ranging from residues 221 to 234, is phosphorylated by an L protein-associated kinase[31] (phosphorylation sites are shown in red in Fig. 7a) and is involved in the binding of P to the L protein.[13,17] Two regions of P were identified as binding sites for the polymerase, one in the N-terminal half of P[15] and another in the C-terminal domain.[13,17] The L binding site in P_{CTD} was assigned to a region of 14 amino

acids ranging from residue 231 to residue 244[17] (shown in green in Fig. 7a). Phosphorylation of VSV P is carried out by two protein kinases acting sequentially.[32] The casein kinase II regulates transcriptional activity by phosphorylating residues in the N-terminal half of P, whereas an L-polymerase-associated kinase phosphorylates additional sites in P_{CTD}. In the IND serotype, important phosphorylation sites in P_{CTD} were mapped to Ser226, Ser227, and Ser233[18,19] (shown in red in Fig. 7a), but the role of these secondary phosphorylations is unclear. It has been proposed that hyperphosphorylation of P could lead to the arrest of RNA replication. These residues precede or are part of the L binding site and could thus directly affect the binding to L. Finally, deletions in the N-terminal part of P_{CTD} affect virus assembly (amino acids 191 to 200, shown in yellow in Fig. 7a) and suppress RNA replication (amino acids 201 to 210, shown in orange in Fig. 7a) without preventing binding to N and L.[12]

The identification of functional regions in P_{CTD} relied mainly on deletion studies. P_{CTD} is a small domain of 71 residues, and thus, truncation of large fragments might have major impacts on its structure. Mutants with large deletions in P_{CTD}, however, conserve their ability to bind to their partners, suggesting that binding occurs through recognition of sequence patterns in P_{CTD} rather than through recognition of 3D motifs. Different parts of VSV P_{CTD}, including the C-terminal positively charged region, are highly flexible as seen by NMR spectroscopy. This feature of P_{CTD} might explain our failure to find crystallization conditions through the use of different crystallization screens (see Materials and Methods) and might be important for the binding to its partners.

Structure conservation within the *Rhabdoviridae*

VSV and RV belong to two different genera of *Rhabdoviridae*, the *Vesiculoviruses* and the *Lyssaviruses*, but multiple sequence alignments indicate almost no sequence similarity between these two P proteins. However, P proteins from these two viruses fulfill similar functions and seem to share a common modular organization. The C-terminal domain in VSV P is shorter than the C-terminal domain of RV P, but the structure of VSV P_{CTD} determined here by NMR spectroscopy is related to that of RV P. The topology of the backbone is conserved, but VSV P_{CTD} lacks helices α_3 and α_6 of RV P_{CTD} (Fig. 5). This new structure suggests that the C-terminal domain and its associated biological functions are conserved among the *Rhabdoviridae*.

The most conserved residues identified through multiple amino acid sequence alignments of all members of the *Vesiculoviruses*, including Isfahan, Piry, and Chandipura viruses, were found in the main hydrophobic core of P_{CTD}, with the largest number of conserved residues located in helix α_3. Restricting sequence comparison to the different VSV strains of both IND and NJ serotypes revealed additional conserved residues mainly in helices α_2 and α_4. On the surface of the protein, the conserved residues are located in the C-terminal part of helix α_3, in helix α_4, and in the short C-terminal loop. They form an elongated patch along one ridge of the molecule and line two large cavities (Fig. 7b). In RV P_{CTD}, most of the residues conserved among *Lyssaviruses* map on the rounded face of the domain (Fig. S4a).

Residues of P_{CTD} from Mokola virus involved in the binding to N were identified by mutagenesis using a two-hybrid system.[33] P protein from Mokola virus is homologous to RV P, and mapping of these residues on the surface of RV P_{CTD} revealed the presence of two major clusters on opposite faces of the domain.[20] The distribution of electrostatic potentials on the surface of RV P_{CTD} revealed a bipolar distribution of charges with a positive pole and a negative pole[20] (Fig. S4b). The first cluster is located in the positive pole of the rounded face and comprises conserved positively charged residues (K211, K212, and K214) and Leu224. The second cluster, located on the opposite flat face of P_{CTD}, forms a hole on the surface of the protein and comprises Cys261, Trp265, and Met287 (Figs. 8a and Fig. S5a). For VSV P_{CTD}, analysis of the distribution of electrostatic potentials (Fig. 7c) also shows a bipolar distribution of charges on the surface of VSV

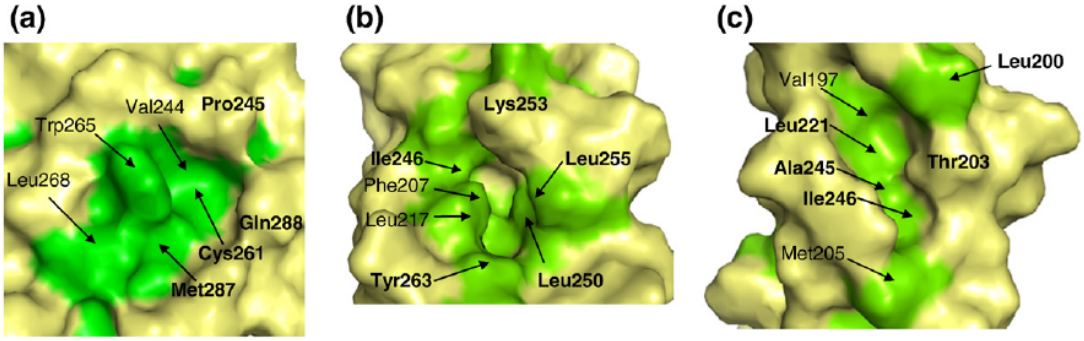


Fig. 8. View of the conserved hydrophobic clusters in RV P_{CTD} (a) and VSV P_{CTD} (b and c). Hydrophobic side chains are shown in green and are labeled. The bold labels highlight the residues conserved in *Lyssaviruses* (a) or in VSV strains (b and c).

P_{CTD} with a pole of positive charges and a pole containing equivalent numbers of positive and negative charges. The positively charged side chains of Lys210, Lys211, Arg251, Lys253, Lys254, Arg260, and Lys262 form a cluster on one face of P_{CTD}. Among these basic residues, Lys210, Lys253, Lys254, Arg260, and Lys262 are strictly conserved among the different VSV strains, while Lys211 and Arg243 found in IND serotypes are replaced by Arg and Lys residues in NJ serotypes, respectively. Arg260 is the only basic residue that is strictly conserved among all *Vesiculoviruses*.

Two deep hydrophobic cavities are present on the surface of VSV P_{CTD} and comprise conserved residues. The first cavity is located between helices α_3 and α_4 and the β-hairpin and is lined by numerous hydrophobic residues (Phe207, Leu217, Ile246, Leu250, Tyr263, and Leu265), of which some are conserved among VSV strains (Figs. 8b and Fig. S5b). This large pocket (1.2 nm long and 0.6 nm wide) is located on a different face of P_{CTD} compared to the cluster of basic residues and might correspond to the hydrophobic cluster found on the surface of RV P_{CTD} (Figs. 8a and Fig. S5a). The second cavity in VSV P_{CTD} is a long groove (1.1 nm long and 0.5 nm wide) located between helix α_1 and the loop connecting helices α_2 and α_3. Its bottom comprises hydrophobic and conserved residues (Figs. 8c and Fig. S5c). These deep cavities at the surface of P_{CTD} might represent binding sites for viral or cellular partners. Deletion studies revealed that three amino acids at the C-terminal extremity of N, including a valine and a phenylalanine, are critical for the association with P.[14] Alternatively, cavities at the surface of P_{CTD} might result from the intrinsic flexibility of this protein domain and indicate that binding to partners involves significant conformational changes in P_{CTD}.

The presence of clusters of conserved basic and hydrophobic residues on opposite faces of VSV and RV P_{CTD} supports a conservation of functional sites on the surface of the protein. It will now be interesting to compare the structure of the central oligomerization domain of RV with that determined for VSV[10] and also to characterize the putative structure of the N-terminal domain of both proteins alone or in complex with N^0 to find out whether the different domains were conserved through evolution. Understanding the origin and evolution of RNA viruses is complicated by the high rate of mutation and specific adaptation to host cells.[34–37] The comparison of genome or protein sequences can only probe closely related viruses, and a more thorough analysis awaits accumulation of structural information on both viral and cellular proteins.

Materials and Methods

Amino acid sequence analysis

The location of disordered regions within VSV P was predicted by submitting its amino acid sequence to 16 different algorithms accessible through web servers. FoldIndex was run with a window size of 51.[38] With the different PONDR predictors (XL3, VL-XT, XL1-XT, and VSL1), residues with a score higher than 0.5 were considered as disordered.[39,40] The NORSp server was used with a window size for disorder set to 25 residues, the structure content cutoff set to 12%, and the number of consecutive exposed residues set to 10.[41] The IUPred predictor for long disordered regions was used and residues with a score above 0.5 were considered as disordered.[42] The DISOPRED server was run with the rate threshold for predictions of false positive set at 2%.[43] The DisEMBL server (Loops/Coils, Hot-loops, and Remark465),[44] RONN server,[45] and PreLink server[46] were run using the default parameters. The SEG program was used for analyzing sequence complexity.[47] Hydrophobic cluster analysis plots were drawn with DRAWHCA,[48] and disordered regions of more than 20 residues were identified by visual inspection. A simple scoring procedure was used to define consensus disordered regions. For each prediction, residues were simply ranked in a binary manner as ordered or disordered using the default threshold set for each algorithm in the web servers. The statistical weights yielded by some of these algorithms are not accounted for, since they are not available for each predicting method and are difficult to compare. For each algorithm, a score of 0 is attributed to each residue predicted to be disordered, whereas a score of 1 is attributed to each one predicted to be in a structured region. The values for each residue are added and divided by the number of used algorithms in order to normalize the score. Consensus disordered regions were defined as regions with a normalized score ≤ 0.50.

Cloning, production, and purification of VSV P and its C-terminal domain

Three constructs of the C-terminal domain of VSV P, P_{CTD1} (amino acids 215–265), P_{CTD2} (amino acids 195–265), and P_{CTD3} (amino acids 185–265) (Fig. 1c), were obtained by PCR amplification from the full-length gene. For each construct, an NdeI restriction site and a methionine codon were introduced downstream of the coding sequence and an XhoI restriction site was introduced upstream. The PCR products were cloned into pET22b between the NdeI and XhoI sites, thus introducing a short linker of two amino acids (Leu-Glu) and a His_6-tag at the C-terminal extremity of the coding sequence. The constructions were verified by DNA sequencing. The proteins were expressed in *E. coli* BL21(DE3). ^{15}N and ^{13}C protein samples for NMR spectroscopy were produced in M9 minimum medium supplemented with 1.0 g L^{-1} of $^{15}NH_4Cl$, 2.0 g L^{-1} of ^{13}C glucose, and MEM vitamins (Gibco).

Soluble proteins were purified in 20 mM Tris–HCl and 150 mM NaCl, pH 7.5 (Buffer A), in a two-step procedure using a Ni^{2+} resin column and a SEC column (Superdex 75 HR 10/30, GE Healthcare) as previously described for full-length VSV P.[11] Inclusion bodies were washed three times by suspending the pellet in Buffer A and centrifuging at 27,200*g* for 30 min. The pellet was then solubilized with 20 mM Tris–HCl, pH 7.5, containing 8 M urea (Buffer B), and the solution was centrifuged at 48,000*g* for 30 min to remove insoluble material. The supernatant was loaded onto Ni^{2+} resin column, pre-equilibrated with Buffer B. The resin was washed successively with 1 bed volume of Buffer B, with 3 bed volumes of 20 mM Tris–HCl at pH 7.5 to refold the protein, with 3 bed volumes of 20 mM Tris–HCl and 1 M NaCl at pH 7.5 supplemented with 10 mM imidazole,

and, finally, with 3 bed volumes of 20 mM Tris–HCl at pH 7.5. The protein was eluted using 20 mM Tris–HCl and 400 mM imidazole at pH 7.5 and loaded onto a SEC column (Superdex 75 HR 10/30, Amersham Pharmacia) equilibrated with 20 mM Tris–HCl, pH 7.5. The separation was performed at a flow rate of 0.5 mL min^{-1}.

The protein preparations were checked by SDS-PAGE, by electrospray mass spectrometry, and by N-terminal sequencing. Mass spectrometry experiments were performed on a Quattro II Mass Spectrometer (Micromass, Altrincham, UK) by continuously injecting the sample with a Type 22 pump from Harvard Apparatus at a flow rate of 5 μL min^{-1}. Proteins were diluted in H_2O containing 1% formic acid at a concentration of about 1 mg mL^{-1}. N-terminal sequencing was performed by Edman degradation on a Procise Model 492 protein sequencer (Applied Biosystems). All protein concentrations were measured by absorbance spectroscopy using Edelhoch's method[49] and the calculated extinction coefficient (ε_{280}=9540 M^{-1} cm^{-1}).

Solutions of purified P_{CTD2} at different concentrations ranging from 4.0 to 19.0 mg mL^{-1} were submitted to extensive crystallization trials with a PixSys4200 robot (Cartesian) using crystal screens I and II and Index (Hampton research) and the vapor diffusion method in drops of 0.1 μl protein solution and 0.1 μl reservoir solution, but no protein crystals were produced.

CD spectroscopy

Far-UV CD spectra were recorded at 20 °C on a JASCO model J-810 CD spectropolarimeter equipped with a Peltier temperature controller. VSV P_{CTD} was diluted to final concentrations of about 40 μM, in 20 mM Tris–HCl, pH 7.5, containing 150 mM NaCl. Spectra were measured in a cuvette with a path length of 1 mm. Mean residue ellipticities were calculated using the number of residues [P_{CTD2}: 71 residues + linker (2 aa) + His-tag (6 aa); P_{CTD3}: 82 residues + linker (2 aa) + His-tag (6 aa)].

The temperature-induced unfolding and refolding were recorded by CD at 222 nm by varying the temperature between 25 and 60 °C using a scan rate of 60 °C/h. VSV P_{CTD} was diluted in 2 mM Tris–HCl, pH 7.5. Protein concentration was 5 μM and path length was 10 mm.

SEC-MALLS-RI

SEC was performed with a S200 Superdex column (GE Healthcare) equilibrated with 20 mM Tris–HCl and 150 mM NaCl, pH 7.5. Separations were performed at 20 °C with a flow rate of 0.5 mL min^{-1}. Typically, 20 μL of protein solution at concentrations of 5–10 mg mL^{-1} was injected. Online MALLS detection and refractive index measurements were performed with a DAWN-EOS detector (Wyatt Technology Corp., Santa Barbara, CA) using a laser emitting at 690 nm and with a RI2000 detector (Schambeck SFD), respectively. Data were analyzed and weight-averaged molar masses (M_w) were calculated using the ASTRA software (Wyatt Technology Corp.) as described previously.[11]

NC binding assay

The interaction between viral NC and P_{CTD2} was investigated by centrifugation through a glycerol cushion. VSV NC was purified from infected BSR cells as described previously.[50] In this assay, VSV NC was incubated in the presence of P_{CTD2} at 20 °C for 10 min. The final NC concentration was 5 μM in 20 mM Tris–HCl, pH 7.5, containing 50 mM NaCl, and P_{CTD2} was added to NC at a 1:1 N:P ratio. These mixtures (40 μL) were loaded on top of 400 μL of 20 mM Tris–HCl, pH 7.5, containing 15% (v/v) glycerol and centrifuged at 25,000 rpm for 2 h at 20 °C (SW55Ti rotor using Ultra Clear tubes of 0.8 mL). After centrifugation, 40 μL of sample present at the top of the glycerol cushion was recovered and the pellet was suspended in 40 μL of 20 mM Tris–HCl, pH 7.5. These fractions were analyzed by 15% SDS-PAGE. Five microliters of each sample was loaded on the gel, and proteins were detected by silver staining.

NMR spectroscopy

Automated sequential backbone resonance assignment of P_{CTD2} was performed using the Batch protocol recently developed in our laboratory for time-optimized automated backbone resonance assignment. The Batch protocol is based on longitudinal relaxation-optimized BEST-type correlation experiments.[51,52] Sequential connectivity information is extracted from these spectra using COBRA,[53] a correlation-coefficient-based analysis tool. This connectivity information is complemented by additional amino-acid-type information extracted from a HADAMAC spectrum.[54] This automated approach provided assignment for 77% of the non-proline residues using NMR data recorded in 15 h on a 600-MHz spectrometer equipped with a cold probe.

Due to the low sensitivity of some peaks located in dynamic regions of the protein, additional standard 3D HNCACB and CBCA(CO)NH experiments were recorded in order to complete the backbone assignment. Aliphatic carbon and proton side-chain resonances were assigned using 4D HC(CCO)NH-type correlation spectra reconstructed from a set of four 3D projections with (H,C) tilt angles of 0°, ±20°, and 90°. ^{1}H and ^{13}C resonance assignment of aromatic groups was done using standard 2D (HB)CB(CGCDCE)HE and (HB)CB(CGCD)HD spectra. The completeness of the final atom assignment is 83.9% for all atoms and 88.2% for ^{1}H atoms.

NMR distance restraints were derived from 3D ^{15}N-edited NOE spectroscopy (NOESY)–HSQC, 3D aliphatic ^{13}C-edited NOESY–HSQC, 3D aromatic ^{13}C-edited NOESY–HSQC, and 3D methyl-selective NOESY–HSQC experiments.[55] The following NOE mixing times were used: 125 ms (aliphatic and aromatic ^{13}C-edited NOESY–HSQC), 100 ms (^{15}N-edited NOESY–HSQC), and 160 ms (3D methyl-selective NOESY–HSQC).

Residual dipolar couplings (RDCs) were collected on a uniformly ^{13}C, ^{15}N-labeled sample suspended in a liquid-crystalline medium consisting of a mixture of 5% pentaethyleneglycol monododecyl ether ($C_{12}E_5$) and hexanol with a molar ratio of 0.96.[56] $^{1}D_{NH}$, $^{1}D_{C'Ca}$, and $^{1}D_{CaHa}$ couplings were obtained using 3D BEST HNCO or HNCOCA-type experiments. Coupling constants were calculated from the difference between line splitting measured in the isotropic and in the anisotropic sample using the NMRPipe nlinLS fitting routine.[57]

R1, R2(CPMG), and {^{1}H}–^{15}N heteronuclear NOE relaxation experiments were acquired at 600 MHz using standard pulse sequences[58] on P_{CTD2}. For P_{CTD3}, only the {^{1}H}–^{15}N heteronuclear NOE relaxation experiment was acquired. The relaxation-caused magnetization decay was sampled at 0, 100, 200, 400, 600, 800, 1100, 1500, and 1900 ms for longitudinal and at 10, 30, 50, 70, 90, 130, 170, 210, and 250 ms for transverse relaxation.

Standard pulse sequences were taken from the Varian BioPack. All NMR spectra were processed and analyzed using NMRPipe and NMRView.[59]

Structure determination

Structure determination was accomplished with the ATNOS–CANDID NOESY assignment program,[60,61] using the CYANA program to calculate structural ensembles.[62] CYANA was used for seven cycles using only NOEs and dihedral angle restraints derived from the program TALOS.[63] Following this, a second run of ATNOS–CANDID was performed starting from cycle 2 using the structures determined with CYANA. In this case, the program CNS,[64] modified to incorporate the program SCULPTOR to handle RDC constraints, was used.[65] The second run involves a restrained molecular dynamics calculation using the standard CNS force field, with the initial sampling period at 1000 K for 5 ps during which time RDCs are scaled from 0.1% to 100% of their final values. A further sampling period of 0.5 ps precedes a slow cooling to 100 K over a period of 5 ps. Steps of 0.3 fs were used throughout the RDC refinement. All alignment tensor parameters were allowed to float throughout the protocol.[66]

Structure analysis and comparisons

Sequence alignments were made using the CLUSTALW2 server,[67] and structure alignments were made with the SSM program.[27] Electrostatic maps were calculated using the Delphi program.[68] Conservation score was calculated with the AL2CO program.[69] Figures of the structure were prepared with PyMOL[70] and Chimera.[71]

PDB accession number

Coordinates have been deposited in the Research Collaboratory for Structural Bioinformatics PDB with the accession number 2k47.

Acknowledgements

This work was supported by the interdisciplinary program "Maladies Infectieuses Emergentes" from the Centre National de la Recherche Scientifique and by a grant from the French Agence Nationale de la Recherche (ANR) [ANR-07-001-01 (ANRAGE) and ANR-05-JCJC-0077] and Lyonbiopôle. We thank Bertrand Lefèbvre, Denis Daveloos, and Bernard Brasmes [CRSSA (Le Centre de recherche du Service de santé Émile Pardé), Grenoble] for mass spectrometry measurements; Jean-Pierre Andrieu [IBS (Institut de Biologie Structurale), Grenoble] for N-terminal sequencing; Rodolfo Rasia (IBS) for help with RDC analysis; and Ewen Lescop (IBS) for the Batch processing protocols. E.A.R. was supported by postdoctoral fellowships from the University Joseph Fourier and from the ANR program (ANR-06-JCJC-0126-01). F.C.A.G. and C.L. were supported by MENRT (Ministere d'Education National, de la Recherche et de la Technologie) fellowship from the French government.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2008.07.028

References

1. Rodriguez, L. L. (2002). Emergence and re-emergence of vesicular stomatitis in the United States. *Virus Res.* **85**, 211–219.
2. Whelan, S. P., Barr, J. N. & Wertz, G. W. (2004). Transcription and replication of nonsegmented negative-strand RNA viruses. *Curr. Top. Microbiol. Immunol.* **283**, 61–119.
3. Kolakofsky, D., Le Mercier, P., Iseni, F. & Garcin, D. (2004). Viral RNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology*, **318**, 463–473.
4. Emerson, S. U. & Wagner, R. R. (1973). L protein requirement for in vitro RNA synthesis by vesicular stomatitis virus. *J. Virol.* **12**, 1325–1335.
5. Emerson, S. U. & Yu, Y. (1975). Both NS and L proteins are required for in vitro RNA synthesis by vesicular stomatitis virus. *J. Virol.* **15**, 1348–1356.
6. Arnheiter, H., Davis, N. L., Wertz, G., Schubert, M. & Lazzarini, R. A. (1985). Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. *Cell*, **41**, 259–267.
7. Qanungo, K. R., Shaji, D., Mathur, M. & Banerjee, A. K. (2004). Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA. *Proc. Natl. Acad. Sci. USA*, **101**, 5952–5957.
8. Ding, H., Green, T. J. & Luo, M. (2004). Crystallization and preliminary X-ray analysis of a proteinase-K-resistant domain within the phosphoprotein of vesicular stomatitis virus (Indiana). *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2087–2090.
9. Green, T. J., Macpherson, S., Qiu, S., Lebowitz, J., Wertz, G. W. & Luo, M. (2000). Study of the assembly of vesicular stomatitis virus N protein: role of the P protein. *J. Virol.* **74**, 9515–9524.
10. Ding, H., Green, T. J., Lu, S. & Luo, M. (2006). Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J. Virol.* **80**, 2808–2814.
11. Gerard, F., Ribeiro, E., Albertini, A., Zaccai, G., Ebel, C., Ruigrok, R. & Jamin, M. (2007). Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution. *Biochemistry*, **46**, 10328–10338.
12. Das, S. C. & Pattnaik, A. K. (2005). Role of the hypervariable hinge region of phosphoprotein P of vesicular stomatitis virus in viral RNA synthesis and assembly of infectious virus particles. *J. Virol.* **79**, 8101–8112.
13. Gill, D. S., Chattopadhyay, D. & Banerjee, A. K. (1986). Identification of a domain within the phosphoprotein of vesicular stomatitis virus that is essential for transcription in vitro. *Proc. Natl. Acad. Sci. USA*, **83**, 8873–8877.
14. Takacs, A. M., Das, T. & Banerjee, A. K. (1993). Mapping of interacting domains between the nucleocapsid protein and the phosphoprotein of vesicular stomatitis virus by using a two-hybrid system. *Proc. Natl. Acad. Sci. USA*, **90**, 10375–10379.
15. Emerson, S. U. & Schubert, M. (1987). Location of the binding domains for the RNA polymerase L and the

ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus. *Proc. Natl. Acad. Sci. USA*, **84**, 5655–5659.
16. Das, T., Pattnaik, A. K., Takacs, A. M., Li, T., Hwang, L. N. & Banerjee, A. K. (1997). Basic amino acid residues at the carboxy-terminal eleven amino acid region of the phosphoprotein (P) are required for transcription but not for replication of vesicular stomatitis virus genome RNA. *Virology*, **238**, 103–114.
17. Paul, P. R., Chattopadhyay, D. & Banerjee, A. K. (1988). The functional domains of the phosphoprotein (NS) of vesicular stomatitis virus (Indiana serotype). *Virology*, **166**, 350–357.
18. Chen, J. L., Das, T. & Banerjee, A. K. (1997). Phosphorylated states of vesicular stomatitis virus P protein in vitro and in vivo. *Virology*, **228**, 200–212.
19. Gao, Y. & Lenard, J. (1995). Cooperative binding of multimeric phosphoprotein (P) of vesicular stomatitis virus to polymerase (L) and template: pathways of assembly. *J. Virol.* **69**, 7718–7723.
20. Mavrakis, M., McCarthy, A. A., Roche, S., Blondel, D. & Ruigrok, R. W. (2004). Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. *J. Mol. Biol.* **343**, 819–831.
21. Blanchard, L., Tarbouriech, N., Blackledge, M., Timmins, P., Burmeister, W. P., Ruigrok, R. W. & Marion, D. (2004). Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. *Virology*, **319**, 201–211.
22. Johansson, K., Bourhis, J. M., Campanacci, V., Cambillau, C., Canard, B. & Longhi, S. (2003). Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. *J. Biol. Chem.* **278**, 44567–44573.
23. Karlin, D., Ferron, F., Canard, B. & Longhi, S. (2003). Structural disorder and modular organization in Paramyxovirinae N and P. *J. Gen. Virol.* **84**, 3239–3252.
24. Hirel, P. H., Schmitter, M. J., Dessen, P., Fayat, G. & Blanquet, S. (1989). Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl. Acad. Sci. USA*, **86**, 8247–8251.
25. Schanda, P., Forge, V. & Brutscher, B. (2006). HETSOFAST NMR for fast detection of structural compactness and heterogeneity along polypeptide chains. *Magn. Reson. Chem.* **44**, S177–S184; Spec No.
26. Holm, L. & Sander, C. (1996). Mapping the protein universe. *Science*, **273**, 595–603.
27. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2256–2268.
28. Lipari, G. & Szabo, A. (1982). A model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546–4559.
29. Houben, K., Blanchard, L., Blackledge, M. & Marion, D. (2007). Intrinsic dynamics of the partly unstructured PX domain from the Sendai virus RNA polymerase cofactor P. *Biophys. J.* **93**, 2830–2844.
30. Takacs, A. M. & Banerjee, A. K. (1995). Efficient interaction of the vesicular stomatitis virus P protein with the L protein or the N protein in cells expressing the recombinant proteins. *Virology*, **208**, 821–826.
31. Chattopadhyay, D. & Banerjee, A. K. (1987). Phosphorylation within a specific domain of the phosphoprotein of vesicular stomatitis virus regulates transcription in vitro. *Cell*, **49**, 407–414.
32. Barik, S. & Banerjee, A. K. (1992). Sequential phosphorylation of the phosphoprotein of vesicular stomatitis virus by cellular and viral protein kinases is essential for transcription activation. *J. Virol.* **66**, 1109–1118.
33. Jacob, Y., Real, E. & Tordo, N. (2001). Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains. *J. Virol.* **75**, 9613–9622.
34. Benson, S. D., Bamford, J. K., Bamford, D. H. & Burnett, R. M. (2004). Does common architecture reveal a viral lineage spanning all three domains of life? *Mol. Cell*, **16**, 673–685.
35. Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. & VandePol, S. (1982). Rapid evolution of RNA genomes. *Science*, **215**, 1577–1585.
36. Bamford, D. H., Grimes, J. M. & Stuart, D. I. (2005). What does structure tell us about virus evolution? *Curr. Opin. Struct. Biol.* **15**, 655–663.
37. Worobey, M. & Holmes, E. C. (1999). Evolutionary aspects of recombination in RNA viruses. *J. Gen. Virol.* **80**, 2535–2543.
38. Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S. *et al.* (2005). FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*, **21**, 3435–3438.
39. Li, X., Romero, P., Rani, M., Dunker, A. K. & Obradovic, Z. (1999). Predicting protein disorder for N-, C-, and internal regions. *Genome Inform. Ser. Workshop Genome Inform.* **10**, 30–40.
40. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J. & Dunker, A. K. (2001). Sequence complexity of disordered protein. *Proteins: Struct. Funct. Genet.* **42**, 38–48.
41. Liu, J., Tan, H. & Rost, B. (2002). Loopy proteins appear conserved in evolution. *J. Mol. Biol.* **322**, 53–64.
42. Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, **21**, 3433–3434.
43. Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F. & Jones, D. T. (2004). The DISOPRED server for the prediction of protein disorder. *Bioinformatics*, **20**, 2138–2139.
44. Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J. & Russell, R. B. (2003). Protein disorder prediction: implications for structural proteomics. *Structure*, **11**, 1453–1459.
45. Yang, Z. R., Thomson, R., McNeil, P. & Esnouf, R. M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, **21**, 3369–3376.
46. Coeytaux, K. & Poupon, A. (2005). Prediction of unfolded segments in a protein sequence based on amino acid composition. *Bioinformatics*, **21**, 1891–1900.
47. Wooton, J. (1994). Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput. Chem.* **18**, 269–285.
48. Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, B., Chomilier, J. *et al.* (1997). Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell. Mol. Life Sci.* **53**, 621–645.
49. Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry*, **6**, 1948–1954.
50. Iseni, F., Baudin, F., Blondel, D. & Ruigrok, R. W. (2000). Structure of the RNA inside the vesicular stomatitis virus nucleocapsid. *RNA*, **6**, 270–281.

51. Schanda, P. & Brutscher, B. (2007). Hyperdimensional protein NMR spectroscopy in peptide-sequence space. *J. Am. Chem. Soc.* **129**, 11916–11917.
52. Lescop, E., Schanda, P. & Brutscher, B. (2007). A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson.* **187**, 163–169.
53. Schanda, P., Van Melckebeke, H. & Brutscher, B. (2006). Speeding up three-dimensional protein NMR experiments to a few minutes. *J. Am. Chem. Soc.* **128**, 9042–9043.
54. Lescop, E., Rasia, R. & Brutscher, B. (2008). Hadamard amino-acid-type edited NMR experiment for fast protein resonance assignment. *J. Am. Chem. Soc.* **130**, 5014–5015.
55. Van Melckebeke, H., Simorre, J. P. & Brutscher, B. (2004). Amino acid-type edited NMR experiments for methyl–methyl distance measurement in ^{13}C-labeled proteins. *J. Am. Chem. Soc.* **126**, 9584–9591.
56. Rückert, M. & Otting, G. (2000). Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.* **122**, 7793–7797.
57. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
58. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G. *et al.* (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry*, **33**, 5984–6003.
59. Johnson, B. A. & Blevins, R. A. (1994). NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 606–614.
60. Herrmann, T., Guntert, P. & Wuthrich, K. (2002). Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR*, **24**, 171–189.
61. Herrmann, T., Guntert, P. & Wuthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**, 209–227.
62. Guntert, P., Mumenthaler, C. & Wuthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298.
63. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
64. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **54**, 905–921.
65. Hus, J. C., Marion, D. & Blackledge, M. (2000). De novo determination of protein structure by NMR using orientational and long-range order restraints. *J. Mol. Biol.* **298**, 927–936.
66. Sibille, N., Blackledge, M., Brutscher, B., Coves, J. & Bersch, B. (2005). Solution structure of the sulfite reductase flavodoxin-like domain from *Escherichia coli*. *Biochemistry*, **44**, 9086–9095.
67. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H. *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
68. Honig, B. & Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science*, **268**, 1144–1149.
69. Pei, J. & Grishin, N. V. (2001). AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics*, **17**, 700–712.
70. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*. DeLano Scientific, Palo Alto, CA, USA.
71. Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C. & Ferrin, T. E. (2006). Tools for integrated sequence–structure analysis with UCSF Chimera. *BMC Bioinformatics*, **7**, 339.

III. Données supplémentaires: simulations de dynamique moléculaire sur le domaine C-terminal de la phosphoprotéine du VSV et les délétions C-terminales ΔC11 et ΔC21.

A. Méthodes

Les simulations de dynamique moléculaire (MD) ont été réalisées à l'aide du logiciel GROMACS 3.3 et du champ de force GROMOS 53a6 (Oostenbrink, Villa et al. 2004; Oostenbrink, Soares et al. 2005). Les modèles de VSV P_{CTD} (résidus 195-265) et de ses délétions C-terminales ΔC11 (résidus 195-254) et ΔC21 (résidus 195-244) ont été immergés dans des boîtes rectangulaires contenant environ 7700 molécules d'eau SPC/E (Berendsen, Grigera et al. 1987), avec une distance minimum entre le soluté et la boîte de 0.9 nm. La charge négative nette du système a été neutralisée par addition d'ions Na^+ et Cl^- à une concentration finale de 20 mM. Les contributions électrostatiques longues distances ont été calculées par sommation d'Ewald (particle-mesh Ewald, PME) avec un cut-off de 0.9 nm. La géométrie des molécules d'eau a été contrainte grâce à l'algorithme SETTLE (Miyamoto and Kollman 1992). Les longueurs des liaisons covalentes ont été contraintes à l'aide de l'algorithme LINCS (Hess, Bekker et al. 1997). La température (300 K) et la pression (1 atm) ont été maintenues par couplage à un bain de chaleur externe et à un bain de pression isotrope avec des constantes de couplage de 0.1 et 1 ps, respectivement. Après 2500 étapes de minimisation d'énergie, le système a été équilibré pendant 100 ps avec les atomes de protéine restreints. Des simulations de 20 ns ont ensuite été calculées et les trajectoires résultantes analysées avec les routines de GROMACS. Les calculs de dérive conformationnelle (Root Mean Square Deviation, RMSD) et de fluctuations conformationnelles (Root Mean Square Fluctuation, RMSF) ont été réalisés sur les Cα après superposition du squelette protéique à la structure de référence. Les figures représentant des structures extraites des trajectoires ont été réalisées dans PyMOL (DeLano 2002) et les graphiques à l'aide de Sigmaplot (SYSTAT Software, Inc.).

B. Stabilité du domaine C-terminal de la phosphoprotéine du virus de la stomatite vésiculaire et des délétions ΔC11 et ΔC21.

Les simulations de dynamique moléculaire réalisées sur VSV P_{CTD} (résidus 195-265) et ses délétions C-terminales ΔC11 (résidus 195-254) et ΔC21 (résidus 195-244) soulignent d'une part la flexibilité conformationnelle du domaine, et indiquent par ailleurs une probable déstabilisation du repliement global du P_{CTD} lorsque les deux hélices α C-terminales sont délétées, comme le suggèrent les courbes de RMSD représentant la dérive conformationnelle du modèle par rapport à la structure de départ au cours des simulations (Figure 25A). La trajectoire obtenue pour VSV P_{CTD} montre une flexibilité importante dans la région 209-215 (Figure 25B) et un dépliement de l'hélice C-terminale α4 (Figure 25C) qui se produit

spontanément à t ~ 15ns. Ces données sont en accord avec les résultats de dynamique obtenus en RMN, qui indiquent que ces deux régions de la protéine sont en échange conformationnel.

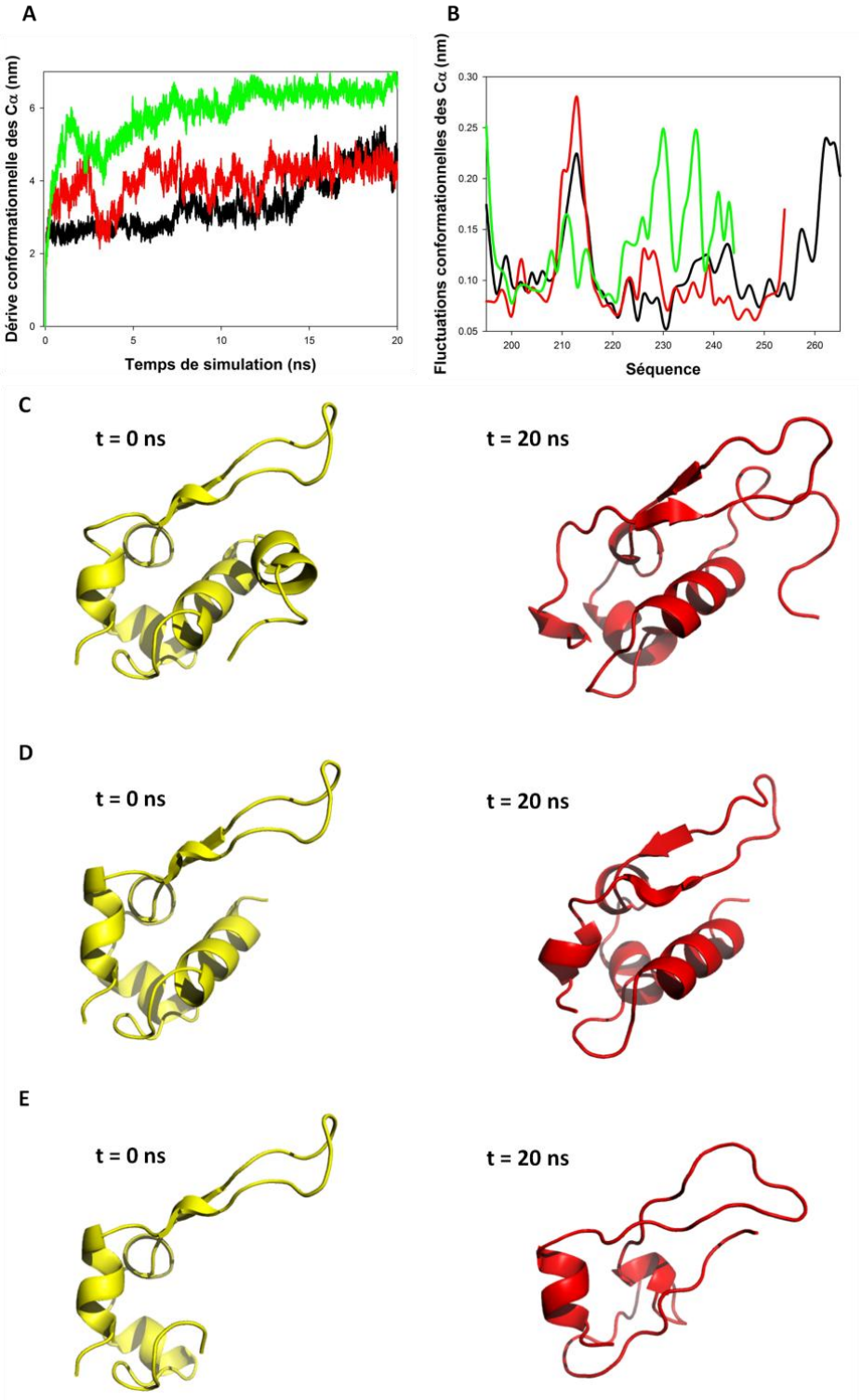


Figure 25 : Simulations de Dynamique Moléculaire de VSV P_{CTD} (résidus 195-265) et de ses délétions C-terminales ΔC11 (résidus 195-254) et ΔC21 (résidus 195-244). A. Dérive conformationnelle des carbones α (RMSD) au cours des simulations de VSV P_{CTD} (noir), ΔC11 (rouge) et ΔC21 (vert). La valeur du RMSD augmente rapidement durant les 5 premières ns et se stabilise en fin de simulation. Une transition est visible à t ~ 15 ns pour VSV P_{CTD}. **B.** Fluctuations conformationnelles moyennées sur les 5 dernières ns de simulation en fonction de la séquence en acides-aminés. Le code couleur est identique à celui utilisé en A. **C, D et E.** Structures initiales et finales des simulations de VSV P_{CTD} (C), ΔC11 (D) et ΔC21(E).

Par ailleurs, l'analyse de la dérive conformationnelle (RMSD, Figure 25A) et des fluctuations de structure en fonction de la séquence (RMSF, Figure 25B) indiquent des changements conformationnels plus importants pour ΔC21 par rapport à ΔC11 et VSV P_{CTD}, ce qui suggère que les 2 derniers éléments de structures hélicoïdales C-terminaux (résidus 245-265) contribuent de manière importante à la stabilité du repliement global de P_{CTD}. En particulier, la structure des brins β1 et β2, (résidus 204-220), déjà déstabilisée par la délétion des 11 derniers résidus dans ΔC11 (Figure 25B&D), ainsi que la structure de l'hélice α2 sont rapidement détruites au cours de la simulation pour la construction ΔC21 (Figure 25B&E).

Ces données théoriques concernant la stabilité de P_{CTD} et de ses délétions C-terminales suggèrent que l'interprétation des données biochimiques basées sur l'utilisation de mutants délétés au niveau de la région C-terminale de P (Paul, Chattopadhyay et al. 1988; Takacs, Das et al. 1993) sont à considérer précautionneusement, le repliement du domaine étant potentiellement rendu trop instable par les mutations.

IV. Conclusion

Les données présentées dans ce chapitre illustrent la validité et l'intérêt des méthodes bioinformatiques utilisées pour définir les différentes régions structurées de la phosphoprotéine, et qui ont été utilisées pour localiser précisément le domaine P_{CTD} du VSV (Chap. II). Plusieurs expériences incluant la dénaturation/renaturation thermique réversible de P_{CTD} par dichroïsme circulaire (CD), la RMN et l'analyse SEC-MALLS de deux constructions correspondant aux résidus 195-265 et 185-265 ont permis de montrer que le domaine P_{CTD} est capable de se replier de manière autonome, et que les résidus 185 à 194 sont effectivement désordonnés, conformément à l'analyse de séquence (Gérard, Ribeiro et al. 2009).

La structure atomique résolue par RMN indique que la topologie du domaine P_{CTD} du VSV est globalement similaire à celle observée pour le virus de la rage (RV). Les différences résident notamment dans la taille du domaine, plus faible pour le VSV, liée à la présence d'une hélice α1 plus courte, et à l'absence des hélices α3 et α6 existant chez RV (Mavrakis, McCarthy et al. 2004).

Les études de relaxation par RMN ont mis en évidence des mouvements rapides (de l'ordre des pico- aux nanosecondes) dans la région de la boucle (209-215) et des échanges conformationnels plus lents dans les hélices N- et C-terminales. Une flexibilité comparativement plus importante a également été observée pour ces régions par Dynamique Moléculaire.

Les similarités topologiques entre les domaines P_{CTD} du VSV et du RV suggèrent que ces deux domaines ont pu évoluer à partir d'un ancêtre commun, une affirmation qui est moins évidente en ce qui concerne l'autre région repliée de la phosphoprotéine, à savoir le domaine central P_{CED}, dont la structure est complètement différente chez VSV et RV (Ding, Green et al. 2006; Ivanov, Crepin et al. 2010).

CHAPITRE V: MODELISATION DE LA PHOSPHOPROTEINE EN SOLUTION

"Consider a spinless ideal Fermi gas of *N* particles. According to Fermi-Dirac statistics, the mean occupation number of a state with energy ε_i is given by

$$\langle n_i \rangle = \frac{1}{e^{(\epsilon_i - \mu)/k_B T} + 1}$$

where, $\langle n_i \rangle$ is the mean occupation number, ε_i is the energy of the i^{th} state, μ is the *fugacity* (which at zero temperature is the maximum energy the particle can have i.e.Fermi energy ε_F)"

(from *Wikipedia* Article on Fermi Surface, 1. Theory)

I. Introduction

Les connaissances accumulées concernant les différentes régions de la phosphoprotéine nous ont permis d'entreprendre la modélisation atomique de cette protéine modulaire en utilisant les données de diffusion de rayons X aux petits angles (SAXS). En particulier, la localisation des régions désordonnées (Gérard, Ribeiro et al. 2009), la localisation des structures α-hélicoïdales résiduelles dans la région N-terminale (Chap.III), la structure cristallographique du domaine de dimérisation (Ding, Green et al. 2006), et la structure RMN en solution du domaine C-terminal (Ribeiro, Favier et al. 2008) ont été utilisées pour générer des ensembles de dimères de phosphoprotéines à l'aide du logiciel Flexible Mecchano développé par Martin Blackledge (Bernado, Blanchard et al. 2005).

La comparaison des résultats obtenus par l'approche d'optimisation d'ensembles par algorithme génétique sur base de la corrélation avec le profil SAXS (Ensemble Optimization Method, (Bernado, Mylonas et al. 2007)) avec l'analyse classique effectuée d'après les données de diffusion de neutrons aux petits angles (SANS) et les paramètres hydrodynamiques de la phosphoprotéine (Gérard, Ribeiro et al. 2007), ainsi qu'avec les méthodes de modélisation *ab initio* (Svergun 1999) ou par assemblage de corps rigides (*i.e. rigid-body modeling*) (Petoukhov and Svergun 2005) nous ont permis d'étendre et de détailler la description structurale de cette protéine multidomaines, en soulignant notamment la flexibilité intrinsèque des dimères de phosphoprotéine en terme de distances et d'orientation relative des régions N- et C-terminales et du domaine central de dimérisation.

II. *ARTICLE V: Modeling the structure of the modular full-length vesicular stomatitis virus phosphoprotein as a conformational ensemble using small-angle X-ray scattering*

Auteurs: Cédric Leyrat, Euripedes de Almeida Ribeiro Jr, Francine Gérard, Malene Ringkjøbing Jensen, Rob Ruigrok, Martin Blackledge and Marc Jamin.

En préparation

Modeling the structure of the modular full-length vesicular stomatitis virus phosphoprotein as a conformational ensemble using small-angle X-ray scattering

Cédric Leyrat[1], Euripedes A. Ribeiro Jr[1], Francine C. A. Gérard[1], Malene Ringkjøbing Jensen[2], Rob W. H. Ruigrok[1], Martin Blackledge[2] and Marc Jamin[1*]

[1] UMI 3265 UJF-EMBL-CNRS, Unit of Virus Host Cell Interactions
6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France
[2]Protein Dynamics and Flexibility, Institut de Biologie Structurale, UMR 5075 CEA-CNRS-UJF, 41 rue Jules Horowitz, Grenoble 38027, France

Keywords: phosphoprotein, intrinsically disordered regions

* **Corresponding author:** Unit of Virus Host Cell Interactions (UVHCI), UMI 3265 UJF-EMBL-CNRS, 6, rue Jules Horowitz, B.P. 181, 38042 Grenoble Cedex 9, France

E-mail : jamin@embl.fr, Phone: + 33 4 76 20 94 62, Fax : + 33 4 76 20 94 00

Abstract

The phosphoprotein of the vesicular stomatitis virus (VSV P) is an essential component of the viral replication machinery that connects the nucleoprotein-RNA (N-RNA) template to the viral polymerase (L) and acts as a chaperone of the nucleoprotein (N) to prevent the non-specific encapsidation of cellular RNAs. VSV P has been previously shown to form elongated dimers in solution (Gérard, Ribeiro et al. 2007) and to possess a modular organization characterized by the presence of two stable, well-folded domains and intrinsically disordered regions (Ribeiro Ede, Leyrat et al. 2009). The high-resolution structure of both central dimerization domain and C-terminal N-RNA binding domain are known. Here, we used and compared different recently developed methods for constructing structural models of the whole VSV P dimer from small-angle scattering data and the known three-dimensional structures. Model-free analysis of the small-angle scattering data and *ab initio* modeling with programs that reconstruct a single model from SAXS data suggested that VSV P dimer has a cylindrical shape with length of 18 nm and diameter of 6 nm. Modeling VSV P as an ensemble of conformers using the ensemble optimization method (EOM) suggested the dynamic character of VSV P dimer and showed that it samples a wide range of interdomain distances and relative orientations. The large range of the radius of gyration (R_g) and of the maximum dimension and (D_{max}) contained in the ensemble proved that the P dimer can continuously extend and compact. These results provide valuable insights into the mechanism of action of P within the viral transcription/replication machineries.

Introduction

The vesicular stomatitis virus (VSV) is a single strand negative-sense RNA virus that served for many years as a model system of the family *Rhabdoviridae*, which includes the prominent rabies virus. The *Rhabdoviridae* family is grouped within the order *Mononegavirales* with three other families of negative-sense RNA viruses, the *Paramyxoviridae* (Sendai virus, measles virus, respiratory syncytial virus, Nipah virus), the *Filoviridae* (Ebola virus, Marburg virus) and the *Bornaviridae* (Borna disease virus), which all share similar genome and structural organizations and similar modes of RNA transcription and replication. The genome of VSV contains a set of five genes that are common to all *MNV* viruses and are organized in the same order along the RNA

genome. They encode successively from the 3' extremity to the 5' extremity the nucleoprotein (N), the phosphoprotein (P), the matrix protein (M), the glycoprotein (G) and the large subunit of the RNA dependent RNA-polymerase (L). The phosphoprotein (P) is an essential component of the transcription/replication machineries that plays multiple roles at different stages of the viral cycle. P forms with nascent N a complex named N^0-P that prevents N to binds to cellular RNA and preserves N in soluble form available for encapsidating newly synthesized RNA genomes and antigenomes. P also acts as a non-catalytic cofactor of the viral polymerase. The RNA genome is encapsidated by N, and the N-RNA complex rather than naked RNA constitutes the actual template of the polymerase complex. However, the L subunit, which carries all enzymatic activities required for RNA synthesis, mRNA capping and methylation and mRNA poly-adenylation, is unable to bind to the N-RNA template. Because P possesses binding site for L in its N-terminal part and for the N-RNA template in its C-terminal domain, it is presumed that P correctly positions L on the template and maintains it attached when the polymerase complex moves along the template.

The modular organization of VSV P was first evidenced by functional studies (Gill, Chattopadhyay et al. 1986; Chattopadhyay and Banerjee 1987; Chattopadhyay and Banerjee 1988; Paul, Chattopadhyay et al. 1988; Das and Banerjee 1993; Takacs, Das et al. 1993; Jacob, Real et al. 2001) and was recently confirmed on a structural basis (Ribeiro, Favier et al. 2008; Gérard, Ribeiro et al. 2009). Using various biophysical methods, we showed that VSV P forms dimers, which are elongated rather than globular (Gérard, Ribeiro et al. 2007). The central dimerization domain (aa 107-177) was isolated by limited proteolysis and its structure was solved by X-ray crystallography (Ding, Green et al. 2004; Ding, Green et al. 2006). A folded C-terminal domain involved in binding the N-RNA template was localized by a meta-prediction of disordered regions (aa 195-265), and its structure was solved by NMR spectroscopy (Ribeiro, Favier et al. 2008). The meta-prediction also predicted a folded domain in the N-terminal 40 amino acids (Gérard, Ribeiro et al. 2009), which contain the binding site for N^0 (Chen, Ogino et al. 2007), but secondary chemical shifts and ^{15}N relaxation NMR experiments indicated that this region is globally disordered and contains two transient α-helical elements (Leyrat et al., 2010, *submitted*). These results suggest that VSV P contains a long intrinsically disordered N-terminal region (aa 1-106) and two autonomous folding domains, the central dimerization domain and the C-terminal N-RNA binding domain separated by a short flexible linker (aa 178-194).

Small-angle scattering of X-ray and neutrons are established methods for characterizing macromolecules and multi-molecular complexes in solution (Koch, Vachette et al. 2003; Putnam, Hammel et al. 2007; Mertens and Svergun 2010). In this study, we performed SAXS experiments with the entire phosphoprotein from VSV and generated structural models using various approaches including *ab initio* modeling, rigid body modeling using the known structure of the domains and atomic-scale structural ensemble description based on the EOM approach.

Results

Small-angle X-ray scattering

Full-length VSV P was produced as a recombinant protein with a C-terminal His-tag. The homogeneity of every sample used in SAXS experiments was checked by size exclusion chromatography combined with detection by multi-angle laser light scattering and refractometry (Wyatt 1998; Gérard, Ribeiro et al. 2007). As previously reported, VSV P eluted as a single peak from the SEC column (Gérard, Ribeiro et al. 2007). The molecular mass determined from light scattering was constant throughout the chromatographic peak and the polydispersity index, M_w/M_n, was lower than 1.01, indicating that VSV P is homogenous with respect to molecular mass and that VSV P forms dimers (Gérard, Ribeiro et al. 2007).

The scattered intensity of full-length VSV P was collected for scattering vector values ($Q = 4\pi(\sin\theta)/\lambda$) ranging from 0.1 to 2.5 nm^{-1} over a concentration range of 1 to 11 $mg.mL^{-1}$. The shape of the profiles was independent of protein concentration and was flat at low Q values indicating the absence of significant aggregation (Figure 1a). Guinier plots were linear for Q values ranging from 0.01 to 0.06 nm^{-1} ($Q.Rg < 1.4$) and showed no significant dependence on protein concentration, confirming the absence of aggregation (Figure 1b). The radius of gyration, R_g,

determined from the Guinier approximation at low Q value was 5.4 nm ± 0.1 nm (Table 1) in agreement with the value of 5.5 nm ± 0.1 nm determined from the pair distribution function (P(r)) (Figure 1c) and with the value of 5.3 ± 0.1 nm previously measured by small-angle neutron scattering (Gérard, Ribeiro et al. 2007). For a solid sphere of the same molecular volume than non-hydrated VSV P, the calculated R_g value (R_0) is 2.1 nm (Tanford 1961), whereas for bovine serum albumin, a protein of 66 kDa, the measured R_g value is 2.9 nm (Kozak 2005). The anisotropy ratio R_g/R_0 of 2.6 – 1.9 and the asymmetrical distribution of interatomic distances exhibiting a maximum length, D_{max}, of 17.0 nm ± 0.5 nm suggested that P dimer has an elongated rather than a globular shape. The plot of ln(I(Q).Q) *versus* Q^2 was linear within satisfactory $Q.R_{xs}$ limits, conforming a elongated molecule and yielding a value of 2.1 ± 0.1 nm for the cross-section (R_{xs}) (Figure 1d and Table 1) (Glatter and Kratky 1982). Assuming a cylindrical shape for P dimer, a length of 17 ± 1 nm and a radius of 3 ± 0.5 nm were calculated from R_g and R_{xs} values (Glatter and Kratky 1982).

Protein concentration (mg/mL)	R_g (nm) (>Guinier)	R_{xs} (nm)	R_g (nm) (>P(r))	D_{max} (nm)
1.0	5.2 ± 0.1	2.0 ± 0.1	5.3 ± 0.2	17.0
5.5	5.4 ± 0.1	2.1 ± 0.1	5.5 ± 0.1	17.0
11.0	5.4 ± 0. 1	2.1 ± 0.1	5.5 ± 0. 1	17.0

a The radius of gyration, R_g, was calculated from the initial slope of the Guinier plot and from the distance distribution function (P(r) *versus* r). The cross sectional radius of gyration, R_{xs}, was calculated from the plot of ln (I(Q) . Q) versus Q^2. Dmax represents the optimal value used in the calculation of the distance distribution function in order to match the Rg value with that obtained from the Guinier plot.

Table 1. Molecular dimensions of VSV P calculated from SAXS experiments

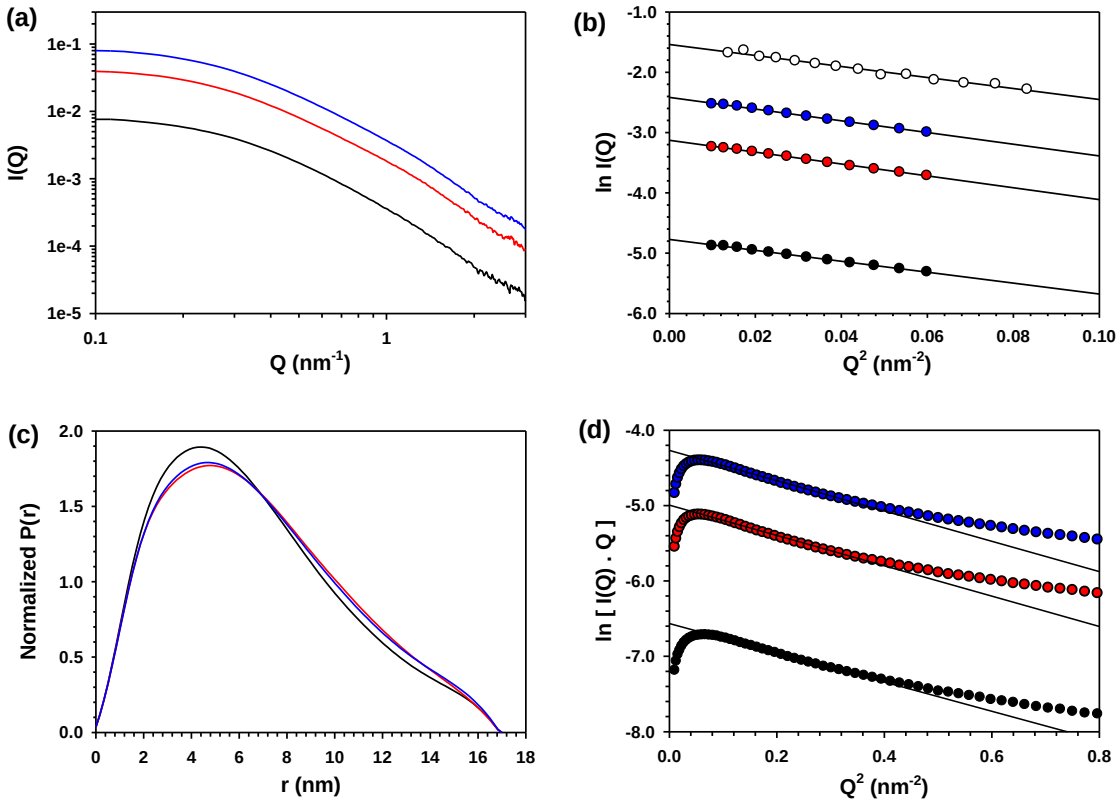


Figure 1. Small-Angle Scattering experiments. A. SAXS data were recorded for Q values in the range 0.14 nm^{-1}< Q < 2.5nm^{-1}. The curves obtained for 3 protein concentrations (3mg/ml, 5.5mg/ml, and 12mg/ml) are represented in black, red, and blue lines, respectively. **B.** Guinier analysis was performed for each three concentrations and the slope is compared to previous SANS measurements (white). **C.** Normalized pair distribution function P(r) at three concentrations.

Ab initio modeling

We set out to build a structural model of full-length VSV P from SAXS data. In a first approach, *ab initio* low resolution bead models were generated with the program DAMMIN (Svergun, 1999). Because the inverse scattering problem has no unique solution, 20 independent models were built and compared. The calculated curve for each models correctly represented the experimental scattering curve with chi values ranging from 0.063 to 0.077. All models had a rugged, cylindrical shape with length of 17 nm corresponding to the maximum dimension of the particle as imposed by default in the program DAMMIN and diameter of about 6 nm in good agreement with triaxial dimensions calculated from R_g and R_{xs} values. The models were aligned pairwise by minimizing the normalized spatial discrepancy (NSD) score with the program DAMAVER. The mean NSD score of 0.75 (values ranging from 0.74 to 0.77) indicated an adequate convergence of the models (Volkov and Svergun 2003). Figure 2 shows three independent models as well as the mean model calculated by the program DAMAVER. Although it was impossible to identify the monomers or to orient the polypeptide chain within the bead models, the structures of dimeric P_{CED} and of two P_{CTD} could easily be accommodated models.

Next, the high resolution structures for P_{CED} and P_{CTD} were used to reconstruct models by rigid body modeling using the program BUNCH (Petoukhov and Svergun 2005). An initial model of a P monomer was built by using the structure of the central and C-terminal domains and by replacing each other amino acid by a densely packed bead. A two-fold symmetry and distance constraints between the central domains of each monomer were imposed in order to reconstitute the dimeric interface. 16 independent models were obtained with the program BUNCH. The agreement between calculated and experimental scattering curves was not as good as that obtained with the program DAMMIN, since chi values ranged from 0.22 to 0.27. All models were elongated in one dimension and had an overall cylindrical shape with length and diameter in agreement with triaxial dimensions calculated from R_g and R_{xs} values, but the NSD values indicated a poor convergence between the different models ($2.38 < NSD < 3.04$).

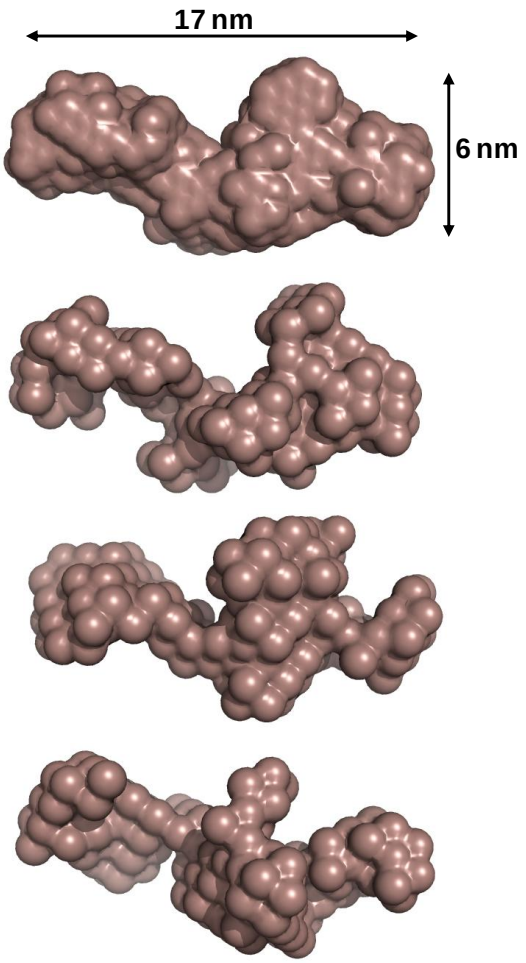


Figure 2. Ab initio models – DAMMIN models. The average model from 20 independent runs is represented as a surface and three views shifted by 90° are shown. The average dimensions of the model are indicated by double arrows.

Rigid body modeling

Visual inspection of the different models revealed two types of organization. In the first type of model, the two P_{CTD}s are close to each other and localized at one end of the cylinder, whereas the two N-terminal regions are localized at the other end. In the second type of model, the N-terminal region of one monomer and the P_{CTD} of the other are close to each other and localized at one extremity of the cylinder.

Modeling P as an ensemble of conformations

Because ^{15}N relaxation measured by NMR spectroscopy showed that the N-terminal part of P and the linker between P_{CED} and P_{CTD} are flexible (Ribeiro, Favier et al. 2008) + (Leyrat et al. 2010, *submitted*), we modeled the structure of P dimer as

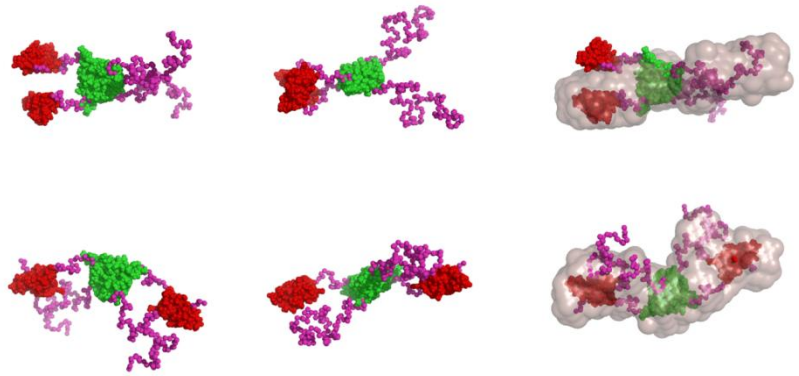


Figure 3. Rigid body models - BUNCH models. Three views shifted by 90° are shown. P_{CTDs} are coloured in red and P_{CED} is represented in green. The remainder of the structure is shown in magenta. The two types of rigid body models (upper and lower panel, respectively) are superimposed onto the average *ab initio* model shown as a semi transparent surface)

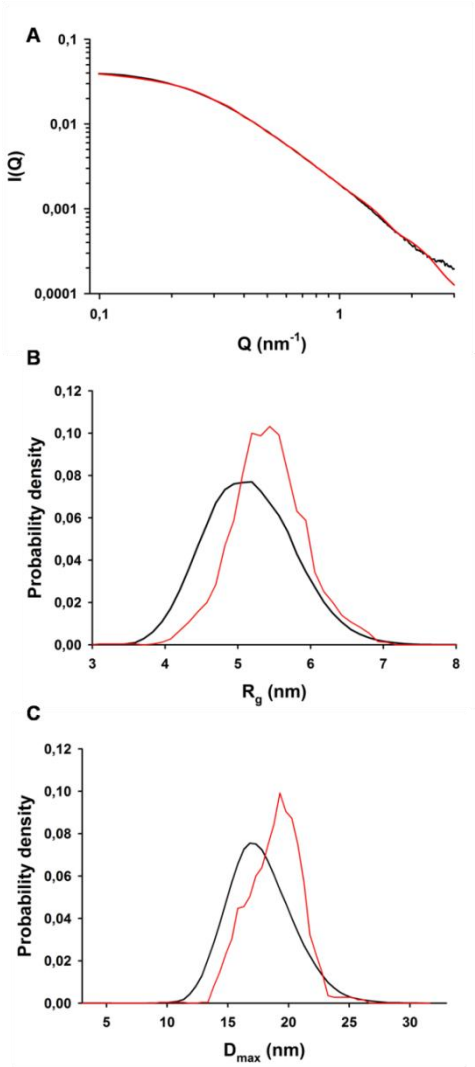


Figure 4. EOM analysis. A. Fitting of SAXS data using the EOM procedure. B. R_g distributions. The black curve shows the R_g distribution calculated for the initial ensemble of conformers, and the red curve shows the R_g distribution of the selected ensemble that fits the experimental SAXS data. C. D_{max} distributions for the initial ensemble of conformers (black) and the selected ensemble (red).

a ensemble of conformers using the ensemble optimization method (EOM) rather than as a single structure. A large ensemble of 8,000 all-atom models was built with the program Flexible Meccano (Bernado, Blanchard et al. 2005) that adequately samples the space with feasible initial models. In each model, the structures of dimeric P_{CED} and of the two P_{CTD} were conserved. Helices of different lengths were introduced in the N-terminal regions (helix 1: aa 2-9 or 2-19, helix 2: aa 25-32 or 25-38) by pooling eight sub-ensembles of 1,000 conformers each as described in Materials and Methods. The remaining part of the N-terminal region and the linker between P_{CED} and P_{CTD} (aa 178-194) of each monomer were assigned random coil conformations. Only a simple exclusion procedure was applied to avoid clashes within and between the chains. From this ensemble of randomly generated conformers, the program EOM selected a sub-ensemble of 50 conformers whose calculated average curve reproduces the experimental SAXS curve up to Q = 3.0 nm^{-1} (Figure 4 A). The χ value of 0.249 indicated a good agreement between the theoretical and the experimental curves. The selected models thus portray possible structures adopted by the protein in solution and collectively represent VSV P dimer (Bernado, Mylonas et al. 2007). The selected ensemble displays a wide range of R_g values comprised between 4 nm and 7 nm, and D_{max} values comprised between 14 nm and 26 nm (Figure 4 B and C). The R_g distribution for the selected models is broad and centered on the experimental R_g value. However, the experimental R_g value is larger than the average R_g value of the initial ensemble (maximum distribution at 5.0 nm), suggesting that, in average, P dimer adopted more elongated

conformations than predicted from random coil statistics. The D_{max} distribution for the selected ensemble also showed that EOM selected elongated models, but the D_{max} distribution of the selected models is not centered on the D_{max} value of 17 nm extracted from the distance distribution function. This discrepancy likely arises from the fact that dimeric P exists as an ensemble of conformations, a phenomenon which causes the indirect transformation from SAS data not to correctly reproduce the longest dimension of the molecule and that has been described elsewhere (Heller 2004; Bernado 2010). The superposition of models from the selected ensemble using the invariant P_{CED} (Figure 5) revealed that the P dimer samples a large conformational space in solution. Although not clearly visible in individual models, the superposition of multiple conformers (Figure 5 B) conveys an impression of a rod-like particle. The P_{CTD} of both protomers occupy adjacent volumes of space rather above the C-terminal end of P_{CED}, forming one end of the rod (Figure 5 A). The orientation and position in space of these two domains are restricted because of the short length of the linker (aa 178-194). The N-terminal 'arms' are much longer and extend further away from P_{CED} in a opposite direction than P_{CTD} or fold back on the side of P_{CED} (Figure 5 C).

In order to determine whether motions of the flexible part are restricted or concerted, we tested for the existence of correlations between various distances and orientation variables. In that purpose, we used two-dimensional plots and compared the distribution of the initial ensemble with that of the selected ensemble. Four representative correlation plots are shown in Figure 6.

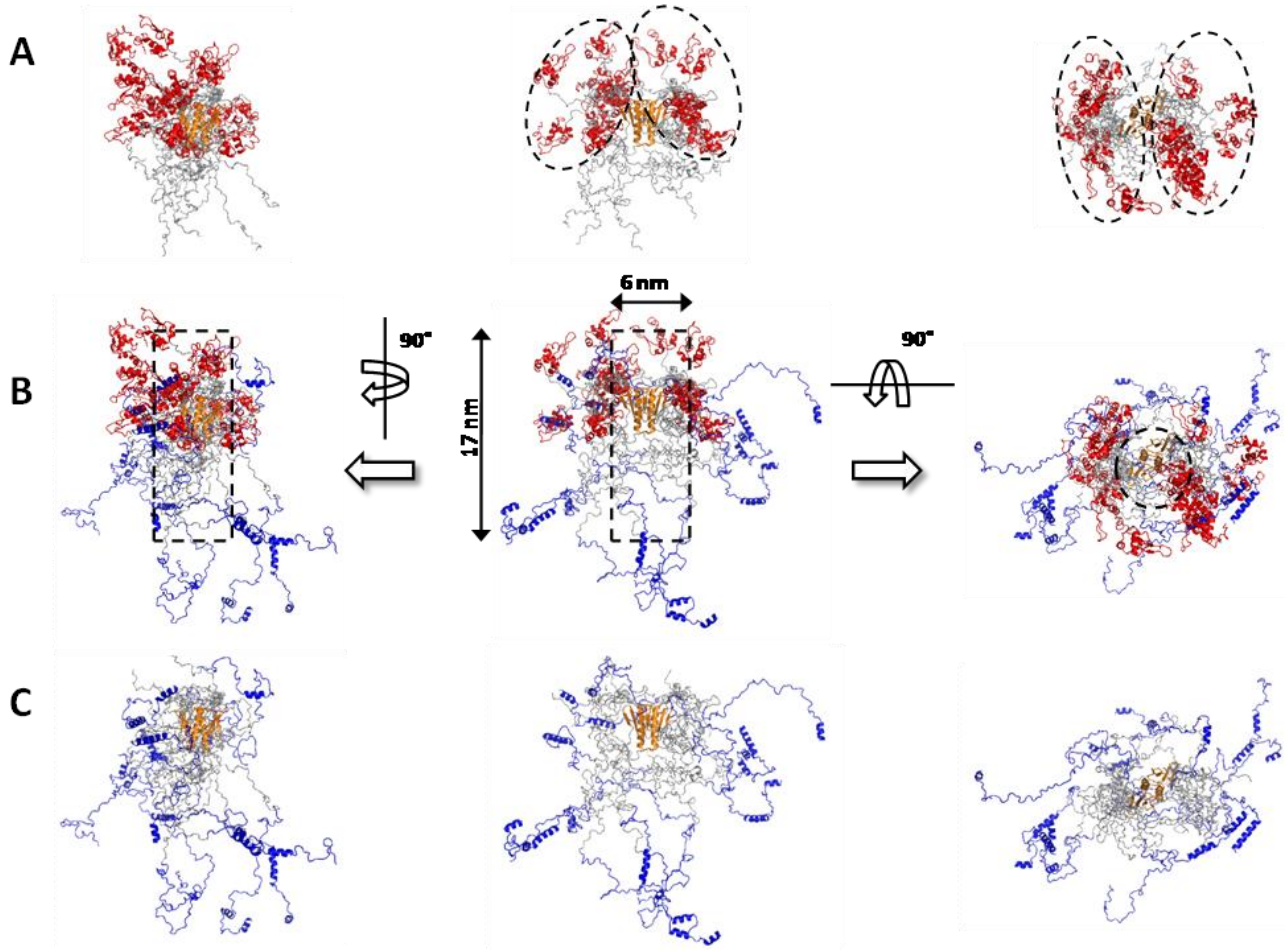


Figure 5. Structural ensemble of VSV P. Five structures from the selected ensemble (50 models) are superimposed onto their respective P_{CED}. P_{CTDs} are shown in red, P_{NTDs} in blue, P_{CED} in orange and the IDR_{NTDs} and IDR_{CTDs} are in grey. Three views shifted by 90° are shown. In A and C, P_{NTD} and P_{CTD} have been respectively removed for clarity. In panel B, the full structure is drawn.

The correlation between P_{NTDs} and P_{CTDs} distances within the VSV P dimers is represented in Figure 6 A. In the selected ensemble, the distance between the center of mass of the two P_{CTDs} vary between 3 nm and 14 nm and the spacing between P_{NTDs} is comprised between 5 and 16 nm. The plot also indicates that these distances are independent. Figure 6 B shows the correlation between the D_{max} of the molecule and the distance separating the P_{CTD} and P_{NTD} from two different monomers of the P protein dimer.

The plot suggests that the selected ensemble is enriched in models with both smaller and larger CTD-NTD distances compared to the average distance found in the pool ensemble. We then calculated correlations between P_{NTDs} and P_{CTDs} spacing angles relative to P_{CED} (Figure 6 C) and the distribution found in the selected ensemble indeed suggests that these angles are independent and can adopt a full range of values (0 – 180°). The relative orientation of the two P_{CTDs} with respect to the P_{CED} was also calculated and the correlation between these angles is shown in Figure 6 D. Similarly to the NTDs and CTDs spacing angles, the relative orientations between P_{CTDs} are evenly distributed. Together, these results indicate that the P_{CTDs} and P_{NTDs} do not suffer from any orientation restrictions within the entire phosphoprotein, suggesting that P flexibility enables it to modulate

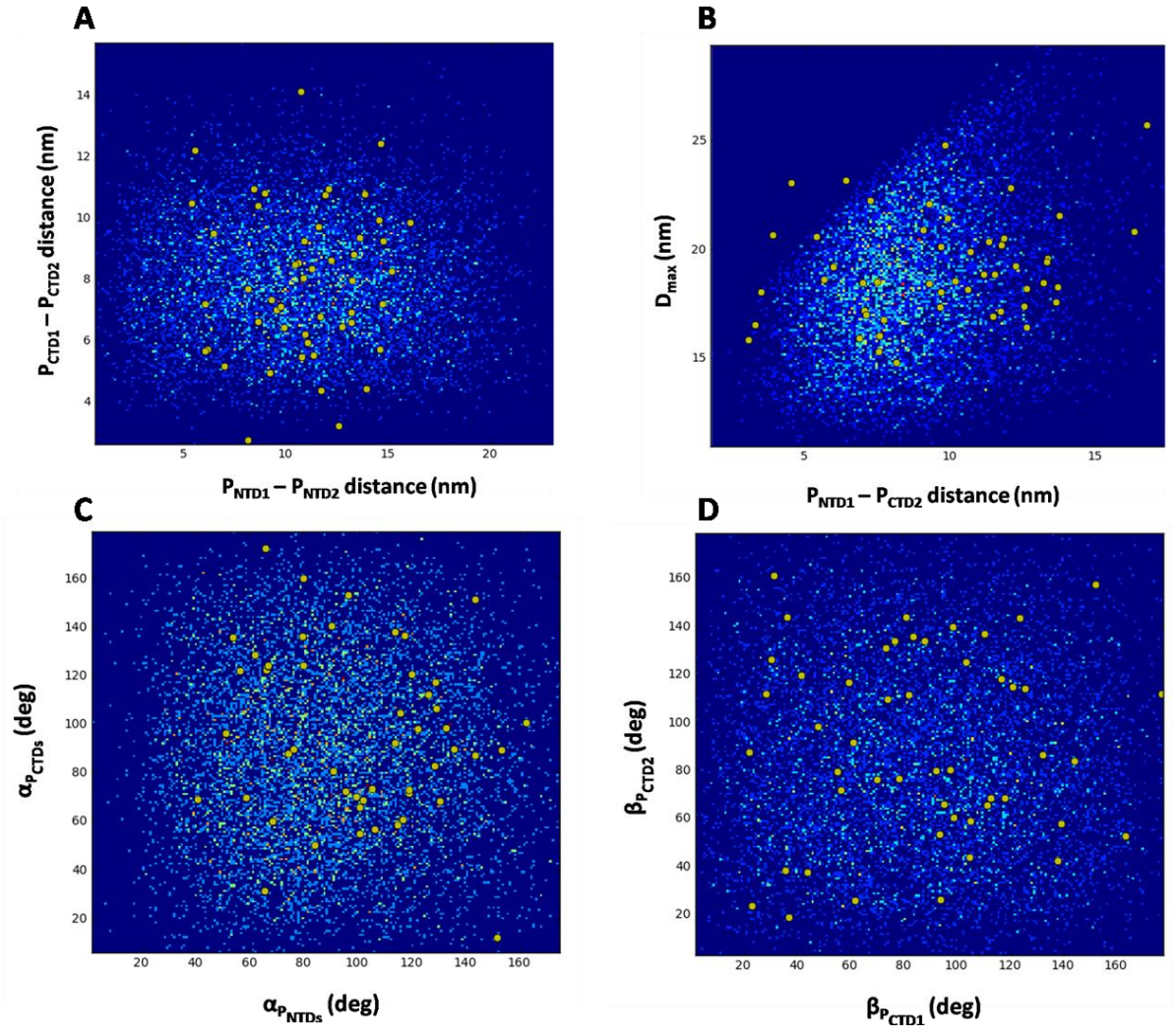


Figure 6. Correlation plots. The distributions of variables (distances or angles) for the initial ensemble of conformers are shown as two-dimensional histograms (blue color indicate no model populates a given region of the plot, while yellow and red colors reflect increased density of population). The range of values displayed by the optimized ensemble of 50 conformers are shown as green-filled circles and superimposed onto the two-dimensional histograms. A. Correlation between P_{NTD1}-P_{NTD2} distance and P_{CTD1}-P_{CTD2} distance. B. Correlation between P_{NTD1}-P_{CTD2} distance and D_{max} of the molecule. C. Correlations correlations between P_{NTDs} and P_{CTDs} spacing angles relative to P_{CED} (α_{PNTDs} and α_{PCTDs}) D. Correlations between β_{PCTD1} and β_{PCTD2}, which represent the relative orientation of each P_{CTD} with respect to the P_{CED}.

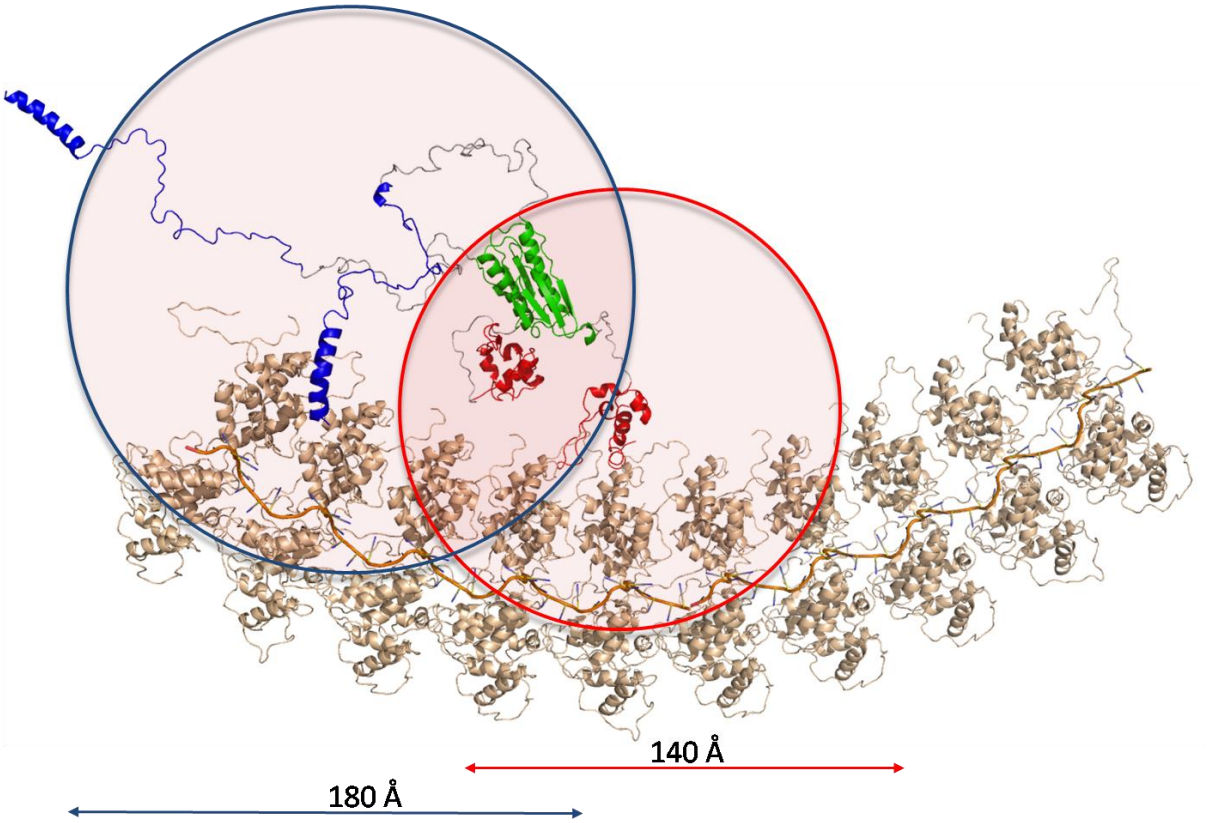


Figure 7. Scheme showing the distance P can reach when attached to N-RNA template. Structural alignments between the N-RNA template structure fitted into Cry-EM density (Ge, Tsao et al. 2010), the crystal structure of P_{CTD} bound to N-RNA (Green and Luo 2009), and a representative model of VSV P (this study) were performed to produce this cartoon illustration of the N-RNA bound phosphoprotein. The color coding scheme used for VSV P is similar to Figure 5, except that P_{CED} is shown in green, and the N-RNA template is shown in wheat. The diameter of the red and blue circles represents the maximum distance observed between the two P_{CTDs}, and between P_{CTD} and P_{NTD}, respectively.

its conformation in order to allow efficient binding to its multiple viral and cellular partners.

Discussion

The different approaches used to interpret SAXS data generated different low resolution models for VSV P dimer. The direct, model-free analysis of SAXS data suggested that VSV P dimer has structural properties of a rod-like particle, *i.e.* an asymmetrical particle, elongated in one direction and with a constant cross-section. As found for such particles, the cross-sectional plot, ln(I(Q) . Q) *versus* Q^2, exhibited a linear portion for Q values ranging from 0.28 to 0.62 nm-1, from which the cross-sectional radius of gyration, R_{xs}, was calculated (Glatter and Kratky 1982). From the experimental R_g and R_{xs} values, VSV P dimer could then be modeled as a simple spherical cylinder with length of 18 nm and a diameter of 6 nm (Glatter and Kratky 1982).

The low-resolution models generated with the programs DAMMIN and BUNCH also displayed an overall cylindrical shape. Models obtained with DAMMIN are bias towards compact models because the program imposes penalties for discontinuities between the beads and therefore ensures the generation of compact and interconnected structures. DAMMIN models of VSV P have an overall cylindrical shape. They are irregular and slightly bent, but it is impossible to orient the polypeptide chain or to localize the folded domains. A dimeric P_{CED} and two P_{CTD} could be manually fitted within the bead model (data not shown). The models generated with BUNCH also exhibited an overall cylindrical shape, but in which the two protomers were distinguishable and oriented. The inclusion of the high-resolution structures yielded more realistic models in which the structured domains were localized. However, modeling dimers with BUNCH could be achieved only by imposing a two-fold symmetry, which is unlikely for VSV P.

However, the main shortcoming of these approaches is that VSV P is modeled as a single structure, whereas several results clearly indicated that P dimer is a highly flexible molecule that can take numerous different conformations. The meta-prediction of disorder performed suggested the presence of a long disordered region in the N-terminal part of P (aa 40-90) and a shorter one between the central and C-terminal domain (aa 186-196) (Gérard, Ribeiro et al. 2009), whereas NMR spectroscopy showed that the N-terminal 60 first amino acids of VSV P (Leyrat et al., 2010, *submitted*) and the linker between P_{CED} and P_{CTD} (Ribeiro, Favier et al. 2008) are globally disordered. The single structures generated with DAMMIN or BUNCH thus, at best, represent average structures of the most populated conformations.

Modeling of VSV P as a conformational ensemble with EOM certainly constitutes an improvement and provides a more realistic picture of VSV P by conveying its dynamic character. The 50 models of the selected ensemble represent different possible conformations present at a given time in the sample or different conformations adopted by the same molecule over time. However, these structures only represent possible conformations that together reproduce the experimental curve. In the selected ensemble, P dimers adopt very different conformations with dimensions and relative orientations between the N- and C-terminal 'arms' that sample the conformational space occupied by the protein in which the N-terminal regions and the flexible linkers connecting P_{CED} to each P_{CTDs} have random coil properties. The analysis of various correlations between independent parameters of dimension and orientation showed that the N-terminal 'arms' and the two P_{CTD} move independently. In this ensemble approach, the D_{max}, R_g and R_{xs} values derived from experimental data represent average values rather than the actual dimensions of the molecule. Conformers more extended than these average dimensions therefore coexist with more compact conformers.

Another clear indication of the flexibility of the molecule comes from the lack of clear interdomain correlation peaks in the p(r) function, a feature of SAXS data that has been shown to correlate with conformational disorder in multidomain proteins (Bernado 2010), and that contrasts with the many defined peaks typically observed in p(r) functions calculated from individual structures (data not shown).

Importance of flexibility for the functions of VSV P

Functional studies with both VSV and RV P suggested that the protein contains independent functional domains that can fold and accomplish their biological function by themselves (Gill, Chattopadhyay et al. 1986; Chattopadhyay and Banerjee 1987; Chattopadhyay and Banerjee 1988; Paul, Chattopadhyay et al. 1988; Das and Banerjee 1993; Takacs, Das et al. 1993; Jacob, Real et al. 2001). A recent meta-prediction of disorder regions as well as biochemical and structural characterizations of various truncated forms of both VSV and RV P demonstrated that these proteins contain folded domains and long intrinsically disordered regions (Ribeiro, Favier et al. 2008; Gérard, Ribeiro et al. 2009). The highly dynamic picture of VSV P allows us to envision how it acts within the transcription/replication complex. P forms dimers that attach to the N-RNA template through P_{CTD}. At this stage, it is not clear if both P_{CTD} of a P dimer can simultaneously attach on the N-RNA template. Binding of one P_{CTD} involves two adjacent N protomers. The maximal distance separating the two P_{CTD} is around 14 nm and is thus long enough for allowing the second P_{CTD} to reach a neighboring binding site (Figure 7). Once the P dimer is bound to the N-RNA complex, the N-terminal flexible 'arms' of P, including the binding sites for viral N^0 and L, and for host cell kinase and exportin, are displayed at the surface of the NC. Because they are highly flexible, they sample a large search volume and can be used to catch their viral and host cell protein partners and attract them towards the template. The maximal distance between P_{CTD} and P_{NTD} is around 18 nm, yielding an estimate of the search volume accessible to the N-terminal 'arms'.

Acknowledgments

This work was supported by grants from the French ANR (ANR-07-001-01 (ANRAGE)), the FINOVI foundation and Lyonbiopôle. E.A.R. was supported by postdoctoral fellowships from both the ANR and the FINOVI programs. C.L. and F.G. were supported by MENRT fellowships from the French government. M.R.J. received a long-term EMBO fellowship and was supported by Lundbeckfonden, Denmark. We thank the

Partnership for Structural Biology for the excellent structural biology environment.

Materials & Methods

Sample preparation

Vesicular stomatitis virus phosphoprotein was produced and purified as previously described (Gérard, Ribeiro et al. 2007). The quality of each sample was checked by size exclusion chromatography (SEC) combined with detection by multi-angle laser light scattering (MALLS) and refractometry (RI) (Wyatt 1998; Gérard, Ribeiro et al. 2007).

Small-angle X-ray scattering (SAXS) experiments

SAXS data were collected at the European Synchrotron Radiation Facility (Grenoble, France) on beamline ID14-3. The sample-to-detector distance was 1 m and the wavelength of the X-rays was 0.0995 nm. Samples were contained in a 1.9 mm wide quartz capillary. The time of exposure was optimized for reducing radiation damage. Data acquisition was performed at 20 °C. Protein concentrations ranged from 1 to 12 mg.ml^{-1}. Data reduction was performed using the established procedure available at ID14-3, and buffer background runs were subtracted from sample runs. The radius of gyration and forward intensity at zero angle (*I(0)*) were determined with the programs PRIMUS (Konarev, Volkov et al. 2003) according to the Guinier approximation at low *Q* values, in a $Q.R_g$ range up to 1.3:

$$\ln I(Q) = \ln I(0) - \frac{R_g^2 Q^2}{3} \qquad \text{(Eq. 1)}$$

The forward scattering intensity was calibrated using bovine serum albumin and lysozyme as references. The cross-sectional radius of gyration (Rxs) was obtained from the plot of ln(I.Q) versus Q2 according to the following equation (Glatter and Kratky 1982):

$$\ln\left(I(Q).Q\right) = \ln\left(I(Q).Q\right)_{Q\to 0} - \frac{R_{xs}^2 Q^2}{2} \qquad \text{(Eq. 2)}$$

The radius of gyration and pairwise distance distribution function, *P(r)*, were calculated with the program GNOM (Semenyuk and Svergun 1991). The maximum dimension (D_{max}) value was adjusted so that the R_g value obtained from GNOM agreed with that obtained from the Guinier analysis.

Modeling of VSV P

Modeling VSV P as a single conformer. The program DAMMIN was used to generate *ab initio* low-resolution models filled with spheres (dummy atoms) by fitting the calculated scattering curve to the experimental curve (Svergun 1999; Petoukhov and Svergun 2005). This program uses a simulated annealing minimization procedure to find the optimum positions of the dummy atoms, starting from a spherical volume of defined radius. By imposing connectivity constraints, DAMMIN minimizes the interfacial area between the molecule and the solvent and therefore favors compact models. 40 independent runs of DAMMIN were performed with no symmetry restriction.

The program BUNCH (Petoukhov and Svergun 2005) was used to perform rigid-body modeling of the full-length protein using the known structures of VSV P_{CED} and P_{CTD}. The regions of the protein for which no high resolution structure was available were modeled by dummy residues. By a simulated annealing protocol, the optimal position and orientation of the folded domains and the conformation of the flexible parts were obtained by fitting the calculated scattering curve to the experimental curve. 16 independent runs of BUNCH were performed with a two-fold symmetry restriction.

Modeling VSV P as an ensemble. A conformational ensemble of 8,000 conformers was generated with the program Flexible-Meccano (Bernado, Blanchard et al. 2005) by pooling 8 sub-ensembles of 1000 conformers in which different amounts of helical structure were imposed in the putative α-helical regions of P_{NTD} identified by NMR spectroscopy as described in (Leyrat et al., 2010, unpublished data). The models consisted of dimers of phosphoprotein including the crystal structure of the central domain (Ding, Green et al. 2006) and the NMR structure of the C-terminal domain (Ribeiro, Favier et al. 2008). Other regions were modeled as unfolded, consistently with disorder predictions (Gérard, Ribeiro et al. 2009). An optimized ensemble of conformations that agrees with the experimental SAXS data was selected from the large conformational ensemble using the ensemble optimization method (EOM) software (Bernado, Mylonas et al. 2007). The procedure was repeated using 10, 20 or 50 models in the selected ensemble with similar results.

The 8,000 conformers from the initial ensemble and the 50 models from the ensemble selected by the EOM software were converted into trajectories using GROMACS 4.0.7 (van der Spoel, Lindahl et al. 2005) and processed using GROMACS tools to calculate various distances and angles. Several angles calculated in order to estimate the extent of interdomain spacing (α_{PNTD} and α_{PCTD}) and orientation (β_{PCTD}) variability. α_{PNTD} was defined as the angle formed between the vector connecting the center of mass (COM) of P_{CED} (residues 107 to 177) to the COM of P_{NTD1}, and the vector connecting the COM of P_{CED} to the COM of P_{NTD2}. α_{PCTD} was similarly defined as the angle between the vectors connecting the COM of P_{CED} to the COMs of P_{CTD1} and P_{CTD2}, respectively. β_{PCTD1} was defined as the angle formed between the two extremities of P_{CTD} (along the largest dimension of the domain) and the COM of P_{CED} and thus enable us to partly monitor the relative orientation of each of the two P_{CTDs} compared to the fixed P_{CED} (indeed, complete monitoring of these relative orientations would require two angles per P_{CTD}). All angles were back calculated from there measured cosines values for each confromer, thus restraining the observed angular range to 0-180°. Python routines were used to calculate the two-dimensional histograms for each combination of variables.

References

1. Gérard, F., et al., *Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution.* Biochemistry, 2007. **46**: p. 10328-10338.
2. Ribeiro Ede, A., Jr., et al., *Binding of rabies virus polymerase cofactor to recombinant circular nucleoprotein-RNA complexes.* J Mol Biol, 2009. **394**(3): p. 558-75.
3. Ding, H., et al., *Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus.* J Virol, 2006. **80**(6): p. 2808-14.
4. Ding, H., T.J. Green, and M. Luo, *Crystallization and preliminary X-ray analysis of a proteinase-K-resistant domain within the phosphoprotein of*

vesicular stomatitis virus (Indiana). Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 11): p. 2087-90.
5. Ribeiro, E.A., Jr., et al., *Solution Structure of the C-Terminal Nucleoprotein-RNA Binding Domain of the Vesicular Stomatitis Virus Phosphoprotein.* J Mol Biol, 2008. **382**: p. 525-538.
6. Gérard, F.C.A., et al., *Modular organization of rabies virus phosphoprotein.* J. Mol. Biol., 2009. **388**: p. 978-996.
7. Chen, M., T. Ogino, and A.K. Banerjee, *Interaction of vesicular stomatitis virus P and N proteins: Identification of two overlapping domains at the N-terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA.* J Virol, 2007. **81**: p. 13478-13485.
8. Koch, M.H., P. Vachette, and D.I. Svergun, *Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution.* Q Rev Biophys, 2003. **36**(2): p. 147-227.
9. Mertens, H.D. and D.I. Svergun, *Structural characterization of proteins and complexes using small-angle X-ray solution scattering.* J Struct Biol, 2010. **172**(1): p. 128-41.
10. Putnam, C.D., et al., *X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution.* Q Rev Biophys, 2007. **40**(3): p. 191-285.
11. Wyatt, P.J., *Submicrometer Particle Sizing by Multiangle Light Scattering following Fractionation.* J Colloid Interface Sci, 1998. **197**(1): p. 9-20.
12. Konarev, P.V., et al., *PRIMUS: a Windows PC-based system for small-angle scattering data analysis.* J. Appl. Cryst., 2003. **36**: p. 1277-1282.
13. Glatter, O. and O. Kratky, *Small Angle X-ray Scattering*. 1982, London, UK: Academic Press.
14. Semenyuk, A.V. and D. Svergun, *GNOM - a program package for small-angle scattering data processing.* J. Appl. Crystallog., 1991. **24**: p. 537-540.
15. Svergun, D.I., *Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing.* Biophys J, 1999. **76**(6): p. 2879-86.
16. Petoukhov, M.V. and D.I. Svergun, *Global rigid body modeling of macromolecular complexes against small-angle scattering data.* Biophys J, 2005. **89**(2): p. 1237-50.
17. Bernado, P., et al., *A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering.* Proc Natl Acad Sci U S A, 2005. **102**(47): p. 17002-7.
18. Ribeiro, E.A., Jr., et al., *Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein.* J Mol Biol, 2008. **382**(2): p. 525-38.
19. Bernado, P., et al., *Structural characterization of flexible proteins using small-angle X-ray scattering.* J Am Chem Soc, 2007. **129**(17): p. 5656-64.
20. van der Spoel, D., et al., *GROMACS: Fast, Flexible and Free.* J. Comp. Chem., 2005. **26**: p. 1701-1718.
21. Tanford, C., *Physical chemistry of macromolecules*, ed. Wiley. 1961, New York.
22. Kozak, M., *Glucose isomerase from Streptomyces rubiginosus – potential molecular weight standard for small-angle X-ray scattering.* J. Appl. Cryst., 2005. **38**: p. 555-558.
23. Volkov, V.V. and D.I. Svergun, *Uniqueness of ab initio shape determination in small-angle scattering.* J. Appl. Cryst., 2003. **36**: p. 860-864.
24. Bernado, P., et al., *Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering.* J Am Chem Soc, 2007.
25. Bernado, P., *Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering.* Eur Biophys J, 2010. **39**(5): p. 769-80.
26. Heller, W.T., *Influence of multiple well defined conformations on small-angle scattering of proteins in solution.* Acta Crystallographica Section D: Biological Crystallography, 2004. **61**(1): p. 33-44.
27. Jacob, Y., E. Real, and N. Tordo, *Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains.* J Virol, 2001. **75**(20): p. 9613-22.
28. Takacs, A.M., T. Das, and A.K. Banerjee, *Mapping of interacting domains between the nucleocapsid protein and the phosphoprotein of vesicular stomatitis virus by using a two-hybrid system.* Proc Natl Acad Sci U S A, 1993. **90**(21): p. 10375-9.
29. Das, T. and A.K. Banerjee, *Acidic domain of the phosphoprotein (P) of vesicular stomatitis virus differentially interacts with homologous and heterologous nucleocapsid protein (N).* Cell Mol Biol Res, 1993. **39**(2): p. 93-100.
30. Paul, P.R., D. Chattopadhyay, and A.K. Banerjee, *The functional domains of the phosphoprotein (NS) of vesicular stomatitis virus (Indiana serotype).* Virology, 1988. **166**(2): p. 350-7.
31. Chattopadhyay, D. and A.K. Banerjee, *NH2-terminal acidic region of the phosphoprotein of vesicular stomatitis virus can be functionally replaced by tubulin.* Proc Natl Acad Sci U S A, 1988. **85**(21): p. 7977-81.
32. Chattopadhyay, D. and A.K. Banerjee, *Two separate domains within vesicular stomatitis virus phosphoprotein support transcription when added in trans.* Proc Natl Acad Sci U S A, 1987. **84**(24): p. 8932-6.
33. Gill, D.S., D. Chattopadhyay, and A.K. Banerjee, *Identification of a domain within the phosphoprotein of vesicular stomatitis virus that is essential for transcription in vitro.* Proc Natl Acad Sci U S A, 1986. **83**(23): p. 8873-7.
34. Ge, P., et al., *Cryo-EM model of the bullet-shaped vesicular stomatitis virus.* Science, 2010. **327**(5966): p. 689-93.
35. Green, T.J. and M. Luo, *Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P.* Proc Natl Acad Sci U S A, 2009. **106**(28): p. 11713-8.

III. *Données supplémentaires: Analyse par diffusion de rayons X aux petits angles de la phosphoprotéine du Virus de la Rage.*

La phosphoprotéine du RV a également été analysée par diffusion de rayons X aux petits angles (SAXS). Les profiles SAXS sont relativement similaires à ceux observés pour le VSV (Figure 26A). Une augmentation du rayon de giration (R_g) calculé à partir de la courbe expérimentale par analyse de Guinier est néanmoins observé à haute concentration (c > 3 mg/ml). Cette augmentation pourrait résulter d'interactions attractives concentration-dépendantes entre les dimères de phosphoprotéines. La présence d'aggrégats dans l'échantillon est cependant exclue car la valeur de rayon de giration mesurée à faible concentration (R_g = 5.10 nm), et obtenue à partir d'une dilution de l'échantillon concentré, est en accord avec la valeur précédemment mesurée par SANS (Gérard, Ribeiro et al. 2007). Le graphe de Kratky (Figure 26B) et la fonction de distribution de distances interatomiques (P(r), Figure 26C)) possèdent une extension aux valeurs de Q élevées, et une forme différente de la forme classique en cloche qui caractérise les objets globulaires, soulignant le caractère partiellement désordonné de la phosphoprotéine. La modélisation par des méthodes *ab initio* fournit des modèles allongés similaires à ce qui est observé chez la protéine P du VSV (données non présentées). Bien qu'une modélisation atomique telle que celle réalisée pour le VSV n'ait pas été réalisée chez RV, ce type de calculs permettrait une description quantitative de la structure de la phosphoprotéine de la rage et autoriserait une comparaison fine avec les résultats obtenus pour le VSV.

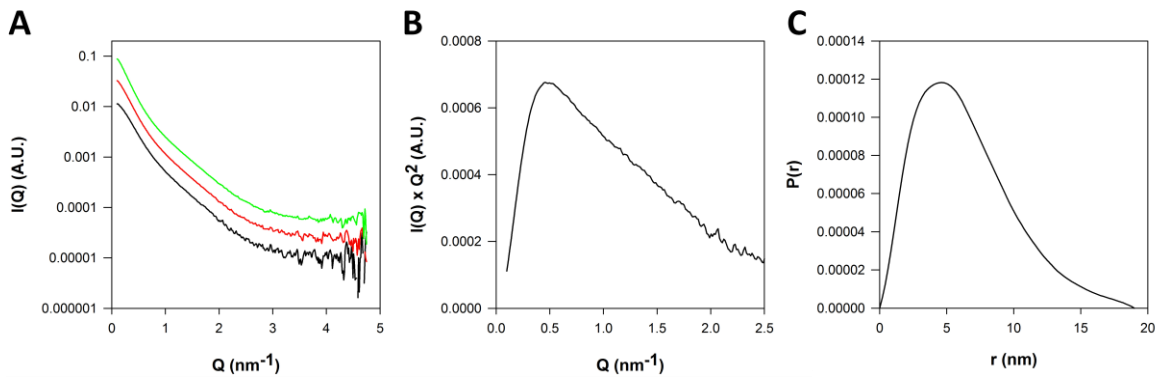


Figure 26 : Expériences de diffusion de rayons X aux petits angles (SAXS) sur la phosphoprotéine du Virus de la Rage. A. Les données de SAXS ont été enregistrées pour des valeurs de Q comprises entre 0.05 nm^{-1}< Q < 4.80 nm^{-1}. Les courbes obtenues à trois concentrations différentes (1, 3, et 7.9 mg/ml) sont représentées en noir, en rouge, et en vert, respectivement. **B.** Graphe de Kratky calculé sur les données à basse concentration (1 mg/ml). **C.** Fonctions de distribution de distances (pair distribution function, P(r)).

IV. Conclusion

L'analyse structurale de la phosphoprotéine par SAXS à l'aide des méthodes *ab initio* ou par ajustement de corps rigides (*« rigid body »*) produit des modèles de la phosphoprotéine aux formes allongées, conformément aux observations basées sur des méthodes hydrodynamiques et sur l'analyse de données de SANS (Gérard, Ribeiro et al. 2007). L'analyse EOM (Ensemble Optimization Method, (Bernado, Mylonas et al. 2007) fournie par GAJOE (Genetic Algorithm Judging the Optimization of Ensembles) suggère une flexibilité importante de la molécule. Aucune restriction angulaire ne semble s'appliquer quant aux orientations relatives des domaines P_{NTD}, P_{CED} et P_{CTD}. L'analyse des distances interdomaines dans le pool de structures partiellement désordonnées et dans l'ensemble sélectionné sur base des données expérimentales indique que la phosphoprotéine est capable de mouvements rapprochant ou éloignant les domaines N- et C-terminaux jusqu'à plus de 18 nm. Ce mouvement joue probablement un rôle important lors de la transcription et de la réplication du virus en permettant notamment le positionnement correct de la polymérase L sur le complexe N-ARN, ou encore en intervenant dans l'acheminement de complexes N^0-P au site de réplication et en catalysant l'encapsidation des génomes et anti-génomes néo-synthétisés.

Par ailleurs, la flexibilité des régions désordonnées (IDR_{CTDs}) joignant les domaines C-terminaux au domaine de dimérisation de la molécule autorise une extension maximale de 14 nm, ce qui suggère que l'association de P avec le complexe N-ARN ne nécessite pas une fixation de P_{CTD} à chacune des sous-unités. La phosphoprotéine semble en effet être en mesure « d'enjamber » plusieurs protomères de nucléoprotéine en utilisant ses deux domaines P_{CTD}.

CHAPITRE VI: INTERACTION DU COMPLEXE NUCLEOPROTEINE-ARN AVEC LA PHOSPHOPROTEINE

"Are you a man..? Or a bag of potatoes?!"

(Euripedes de Almeida Ribeiro Jr, s'adressant à Nicolas Martinez)

I. Introduction

Ce chapitre présente des données concernant l'interaction des complexes N-ARN circulaires avec la phosphoprotéine, avec son domaine P_{CTD}, ainsi qu'avec un mutant $P_{\Delta 91\text{-}133}$ délété de son domaine de dimérisation P_{CED}. Les expériences de titration réalisées par Francine Gérard ont montré que la fixation de deux dimères P_2 par anneau N_{10}-ARN ou N_{11}-ARN était possible, et des expériences de résonance plasmonique de surface (SPR) réalisées par Euripedes de Almeida Ribeiro Jr indiquent que l'affinité du domaine P_{CTD} pour N_{10}-ARN est de 160 ± 20 nM.

Mon implication dans ce projet a consisté principalement à modéliser *in silico* l'interaction entre la phosphoprotéine ou son domaine P_{CTD} et les complexes N_{10}-ARN et N_{11}-ARN à l'aide des structures cristallographiques disponibles (Mavrakis, McCarthy et al. 2004; Albertini, Wernimont et al. 2006; Ivanov, Crepin et al. 2010). Pour ce faire, j'ai utilisé une procédure d'arrimage moléculaire (« *docking* ») flexible adaptée d'après les principes développés dans (Krol, Chaleil et al. 2007; Krol, Tournier et al. 2007), et intégré les différentes données disponibles : les profils de SAXS, la protéolyse limitée (Schoehn, Iseni et al. 2001), la mutagénèse dirigée sur Mokola Virus P_{CTD} identifiant des résidus impliqués dans la fixation à N-ARN et localisés sur les deux faces opposées du domaine P_{CTD} (Jacob, Real et al. 2001), l'exposition de l'épitope 402-13 de P_{CTD} reconnu uniquement sous la forme complexée à N-ARN (Toriumi, Honda et al. 2002), et enfin la compatibilité du modèle sélectionné avec une augmentation d'affinité induite par la phosphorylation du complexe au niveau du résidu Ser389 de la nucléoprotéine (Toriumi and Kawai 2004). J'ai également participé à la production et à la purification de P_{CTD} et de P_2, en compagnie de Francine Gérard et Euripedes de Almeida Ribeiro Jr en vue des expériences de SAXS d'Octobre 2007.

II. ARTICLE VI: Binding of Rabies Virus Polymerase Cofactor to Recombinant Circular Nucleoprotein-RNA Complexes

Auteurs: Euripedes de Almeida Ribeiro Jr*, Cédric Leyrat*, Francine Gérard*, Aurélie Albertini, Caroline Falk, Rob Ruigrok et Marc Jamin.

* Euripedes de Almeida Ribeiro Jr, Cédric Leyrat et Francine Gérard ont contribués de manière équivalente

Publié en 2009 dans *Journal of Molecular Biology*, Volume 394, pages 558-575

doi:10.1016/j.jmb.2009.09.042 J. Mol. Biol. (2009) 394, 558–575

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Binding of Rabies Virus Polymerase Cofactor to Recombinant Circular Nucleoprotein–RNA Complexes

Euripedes de Almeida Ribeiro Jr†, Cédric Leyrat†, Francine C. A. Gérard[1]†, Aurélie A. V. Albertini, Caroline Falk, Rob W. H. Ruigrok and Marc Jamin*

Unit of Virus Host Cell Interactions (UVHCI), UMI 3265 UJF-EMBL-CNRS, 6, rue Jules Horowitz, B.P. 181, 38042 Grenoble Cedex 9, France

Received 16 June 2009; received in revised form 11 September 2009; accepted 16 September 2009
Available online
23 September 2009

In rabies virus, the attachment of the L polymerase (L) to the viral nucleocapsids (NCs)—a nucleoprotein (N)–RNA complex that serves as template for RNA transcription and replication—is mediated by the polymerase cofactor, the phosphoprotein (P). P forms dimers (P_2) that bind through their C-terminal domains (P_{CTD}) to the C-terminal region of the N. Recombinant circular N_m–RNA complexes containing 9 to 12 protomers of N (hereafter, the subscript m denotes the number of N protomers) served here as model systems for studying the binding of P to NC-like N_m–RNA complexes. Titration experiments show that there are only two equivalent and independent binding sites for P dimers on the N_m–RNA rings and that each P dimer binds through a single P_{CTD}. A dissociation constant in the nanomolar range (160±20 nM) was measured by surface plasmon resonance, indicating a strong interaction between the two partners. Small-angle X-ray scattering (SAXS) data and small-angle neutron scattering data showed that binding of two P_{CTD} had almost no effect on the size and shape of the N_m–RNA rings, whereas binding of two P_2 significantly increased the size of the complexes. SAXS data and molecular modeling were used to add flexible loops (N_{NTD} loop, amino acids 105–118; N_{CTD} loop, amino acids 376–397) missing in the recently solved crystal structure of the circular N_{11}–RNA complex and to build a model for the N_{10}–RNA complex. Structural models for the N_m–RNA–$(P_{CTD})_2$ complexes were then built by docking the known P_{CTD} structure onto the completed structures of the circular N_{10}–RNA and N_{11}–RNA complexes. A multiple-stage flexible docking procedure was used to generate decoys, and SAXS and biochemical data were used for filtering the models. In the refined model, the P_{CTD} is bound to the C-terminal top of one N protomer (N_i), with the C-terminal helix (α_6) of P_{CTD} lying on helix α_{14} of N_i. By an induced-fit mechanism, the N_{CTD} loop of the same protomer (N_i) and that of the adjacent one (N_{i-1}) mold around the P_{CTD}, making extensive protein–protein contacts that could explain the strong affinity of P for its

*Corresponding author. E-mail address:jamin@embl.fr.

† E.A.R., C.L., and F.C.A.G. contributed equally to this work.

Present addresses: F. C. A. Gérard, IRCM, Laboratoire de Rétrovirologie Humaine, 110, avenue des Pins Ouest, Montréal, Québec, Canada H2W 1R7; A. A. V. Albertini, CNRS, UMR2472, INRA, UMR1157, IFR 115, Virologie Moléculaire et Structurale, 91198, Gif sur Yvette, France.

Abbreviations used: L, L polymerase; NC, nucleocapsid; N, nucleoprotein; P, phosphoprotein; SAXS, small-angle X-ray scattering; *MNV*, *Mononegavirales*; RV, rabies virus; VSV, vesicular stomatitis virus; MALLS, multi-angle laser light scattering; RI, refractometry; SPR, surface plasmon resonance; SANS, small-angle neutron scattering; SEC, size-exclusion chromatography; EMG, exponentially modified Gaussian; PDB, Protein Data Bank; MD, molecular dynamics; RU, response units.

template. The structural model is in agreement with available biochemical data and provides new insights on the mechanism of attachment of the polymerase complex to the NC template.

Keywords: Rabies virus; Replication complex; Phosphoprotein; Rhabdovirus; molecular modeling

Edited by D. E. Draper

Introduction

Numerous non-segmented negative-strand RNA viruses [*Mononegavirales* (*MNV*) order] cause human and animal diseases. Rabies virus (RV) is a prototypical member of this order that shares with other members a common genome organization and similar transcription/replication mechanisms. The transcription and replication machineries of these viruses involve three viral proteins in addition to the genomic RNA: the nucleoprotein (N) and a two-subunit RNA-dependent RNA polymerase made of the large subunit [L polymerase (L)] and of its cofactor, the phosphoprotein (P).[1,2] N encapsidates the genomic RNA forming long and flexible helical nucleocapsids (NCs). This N–RNA complex serves as a template for both RNA transcription and replication.[3] When expressed in bacteria or insect cells, recombinant N protein binds to cellular RNA and forms circular N_m–RNA complexes in addition to long NCs.[4,5] Circular N_m–RNA complexes containing different numbers of N protomers can be separated by preparative gel electrophoresis and serve here as model systems for studying the interactions with P.[6] The crystal structure of circular N_m–RNA complexes from RV and vesicular stomatitis virus (VSV) revealed that N is a two-domain protein that enwraps completely the viral RNA molecule and hides it from the polymerase and from the innate immune system.[7,8] The L protein has more than 2100 amino acids and carries out all enzymatic activities, including the RNA-dependent RNA polymerase, mRNA capping, and mRNA polyadenylation. P is a multi-domain protein that belongs to the intrinsically disordered proteins[9,10] and plays multiple roles in the viral cycle. P chaperones nascent RNA-free N (N^0), preventing its binding to cellular RNA and delivering N^0 to the site of viral replication for the encapsidation of the newly replicated genomic RNA molecule.[11] The binding site for N^0 is situated in the first 60 N-terminal residues of P,[11] a region that was found to be disordered in isolated P but was predicted to be structured,[9] suggesting that this region may fold upon binding to N^0. For RNA transcription and replication, L uses exclusively genomic RNA molecules enwrapped by Ns, but L is unable, by itself, to attach to the N–RNA template. Moreover, *MNV* polymerases are processive enzymes that must remain attached to the template during transcription and replication.[2,12] P contains a binding site for L in the first 19 N-terminal residues[13] and a binding site for N–RNA in its C-terminal domain (P_{CTD}).[4,14] The binding site for P on L was mapped to the 600 C-terminal amino acids of L[13] and that for P_{CTD} on N–RNA was mapped to the C-terminal region of N (amino acids 377–450) by limited proteolysis.[4,15] However, the detailed mechanisms by which P attaches to both L and N–RNA and ensures the processivity of L remain unknown. Characterizing the structure and properties of a complex between P and the N–RNA template will help to decipher the molecular mechanisms by which these viruses synthesize RNA.

Here, we studied the formation of complexes between different circular N_m–RNA complexes and P dimers, P_{CTD}, or a deletion mutant lacking the central dimerization domain ($P_{\Delta 91-131}$). We used size-exclusion chromatography (SEC) combined with detection by multi-angle laser light scattering (MALLS) and refractometry (RI) to identify different complexes and determine their stoichiometry, and we used surface plasmon resonance (SPR) for measuring the affinity of P dimers for the N_{10}–RNA complex. To characterize the size and shape of the complexes formed between P_{CTD} and the N_m–RNA rings, we measured small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS) profiles, and we turned to molecular modeling, using SAXS and available biochemical data, to select the best models. The refined model of the complex provided new insights on the mechanism of attachment of the two-subunit polymerase complex of RV to its N–RNA template.

Results

Formation of different N_m–RNA–$(P_2)_n$ complexes

Recombinant circular forms of N_m–RNA complexes containing 9, 10, 11, or 12 N protomers were purified by preparative electrophoresis and identified by electron microscopy.[6] Previously, we showed that P_{CTD} is monomeric[16] and that full-length P is dimeric (P_2).[10] The different circular N_m–RNA complexes eluted at different volumes (V_{el}) from a SEC column with Stokes' radii ranging from 6.3 to 6.8 nm (Table 1). Their molecular masses measured by combining SEC with online detection by MALLS and RI (SEC–MALLS–RI)[17] were constant throughout the elution peak, and the polydispersity factors ($\langle M_w \rangle / \langle M_n \rangle$) were close to unity (Table 1), as expected for monodisperse species. The weight-averaged molecular masses ($\langle M_w \rangle$) obtained by averaging measurements throughout the

Table 1. Molecular dimensions of the circular N_m–RNA, N_m–RNA–(P_2), and N–RNA–$(P_2)_2$ complexes determined from SEC, MALLS, and SAXS experiments

N–RNA complex	R_S (nm) (SEC)	R_g (nm) (Guinier)	R_g (nm) [$P(r)$]	D_{max} (nm) [$P(r)$]	MM_{calc} (kDa)	⟨MM⟩ (kDa) (MALLS)	$\langle M_w \rangle / \langle M_n \rangle$ (MALLS)
N_9–RNA	6.3±0.1	6.2±0.1	N.D.	N.D.	483	470±10	1.005±0.040
N_9–RNA–(P_2)	6.8±0.1	N.D.	N.D.	N.D.	550	N.D.	N.D.
N_9–RNA–$(P_2)_2$	7.2±0.1	N.D.	N.D.	N.D.	617	636±10	1.001±0.020
N_{10}–RNA	6.4±0.1	6.1±0.1	5.9±0.2	16.0±0.5	537	510±7	1.001±0.020
N_{10}–RNA–(P_2)	6.9±0.1	N.D.	N.D.	N.D.	604	N.D.	N.D.
N_{10}–RNA–$(P_2)_2$	7.4±0.1	7.0±0.1	7.0±0.2	23.0±0.5	671	666±9	1.000±0.020
N_{10}–RNA–$(P_{\Delta 91-131})$	6.8±0.1	N.D.	N.D.	N.D.	566	N.D.	N.D.
N_{10}–RNA–$(P_{\Delta 91-131})_2$	7.3±0.1	N.D.	N.D.	N.D.	596	565±4	1.000±0.010
N_{10}–RNA–$(P_{CTD})_2$	6.6±0.1	6.1±0.1	6.0±0.1	16.0±0.5	565	532±4	1.000±0.010
N_{11}–RNA	6.7±0.1	6.3±0.1	6.2±0.1	17.0±0.5	590	570±10	1.002±0.025
N_{11}–RNA–(P_2)	7.2±0.1	N.D.	N.D.	N.D.	657	N.D.	N.D.
N_{11}–RNA–$(P_2)_2$	7.6±0.1	7.1±0.2	7.1±0.1	23.0±0.5	724	717±8	1.000±0.015
N_{11}–RNA–$(P_{CTD})_2$	N.D.	6.3±0.1	6.2±0.1	17.0±0.5	618	N.D.	N.D.
N_{12}–RNA	6.8±0.1	N.D.	N.D.	N.D.	644	650±20	1.005±0.050
N_{12}–RNA–(P_2)	7.2±0.1	N.D.	N.D.	N.D.	711	N.D.	N.D.
N_{12}–RNA–$(P_2)_2$	7.6±0.1	N.D.	N.D.	N.D.	778	N.D.	N.D.

N.D., not determined.

chromatographic peak were in agreement with the molecular masses calculated for the N_m–RNA rings from the molecular mass of the N protein (MM of N = 50.7 kDa) and the mean molecular mass of the RNA molecule (average MM of a nucleotide = 330 Da), assuming that each protomer of N bound to nine nucleotides[5] (9 × 330 Da = 3.0 kDa) (Table 1).

We studied the binding of full-length P dimers to the N_m–RNA complexes by mixing the N_m–RNA complexes with increasing amounts of P. The size and stoichiometry of the complexes formed between N_m–RNA complexes and P_2 were determined by SEC–MALLS–RI. The N_{10}–RNA ring eluted at 9.8 ml, while P_2 eluted at 12.2 ml (Fig. 1a). Upon addition of increasing concentrations of P to a fixed concentration of N, the elution peak corresponding to N_{10}–RNA progressively shifted to lower elution volumes (9.8 to 9.1 ml) (Fig. 1a), indicating the formation of N_{10}–RNA–$(P_2)_n$ complexes (hereafter, the subscript n denotes the number of bound P dimers). For [P]/[N_{10}–RNA ring] ratios above 4, the elution volume of the N_m–RNA–$(P_2)_n$ complex remained constant, and a peak of free P_2 became detectable, indicating that under these conditions, rings were saturated with two P_2 molecules. In the concentration range used in these experiments (3 μM < [P] < 300 μM), free P_2 was not detectable until saturation was reached, indicating a K_d lower than 3 μM.

Because the experiments were carried out at concentrations of P above the K_d, the stoichiometry could be estimated from a titration plot, where the elution volume of N_m–RNA–$(P_2)_n$ complex was plotted as a function of the [P]/[N_{10}–RNA ring] molar ratio. Figure 1b shows the titration plot for N_{10}–RNA. The plot exhibits two linear regions that intersects at a [P]/[N_{10}–RNA ring] ratio of 4.4±0.3, suggesting a binding capacity of two dimers of P per ring. At saturation by P, the molecular mass of the N_{10}–RNA–$(P_2)_n$ complex determined by SEC–MALLS–RI was 666±9 kDa, also in agreement with the binding of two P dimers (calculated mass = 537 + 4 × 33.6 = 671 kDa) (Fig. 1c). Similar data and titration plots were obtained with N_9–RNA, N_{11}–RNA, and N_{12}–RNA complexes (Supplementary Fig. S1) and yielded maximum binding capacity values of 4.2±0.1, 4.4±0.2, and 4.0±0.1, respectively. The molecular masses of the different N_m–RNA–$(P_2)_2$ measured by MALLS at saturation with P and the calculated masses for complexes containing one ring and two P_2 are given in Table 1.

At intermediate concentrations of P, the peak width of the N_{10}–RNA–$(P_2)_n$ complex increased significantly, suggesting the existence of multiple species, whereas at high concentrations of P, the peak width of the N_{10}–RNA–$(P_2)_2$ complex was similar to that of the free N_{10}–RNA complex, indicating the presence of a single species. The widening of the chromatographic peak at intermediate concentrations of P and the stoichiometry of four P per N_{10}–RNA ring suggested the formation of complexes containing one or two dimers of P. The chromatographic profiles measured by RI were deconvoluted, assuming the presence of three species under the N_{10}–RNA–$(P_2)_n$ peak, the free N_{10}–RNA complex ($n = 0$), the N_{10}–RNA complex containing one bound dimer of P ($n = 1$), and the one containing two bound dimers of P ($n = 2$) (Fig. 2a). Each species was fitted with a four-parameter exponentially modified Gaussian (EMG) equation. The elution profiles from the titration series (Fig. 1a) were deconvoluted to evaluate the amount of each of the three species at each [P]/[N_{10}–RNA] ratio (Fig. 2b and Supplementary Fig. S2). The titration by SEC occurred in two stages. In the first stage, the N_{10}–RNA–(P_2) complex accumulated at the expense of N_{10}–RNA complex and reached a maximal population at a [P]/[N_{10}–RNA] ratio of about 2. In the second stage, N_{10}–RNA–$(P_2)_2$ accumulated while N_{10}–RNA–(P_2) as well as the remaining N_{10}–RNA disappeared. The titration was correctly simulated by a simple statistical distribution, confirming that the two binding sites for P dimers were independent (Fig. 2b). Similar deconvolutions were

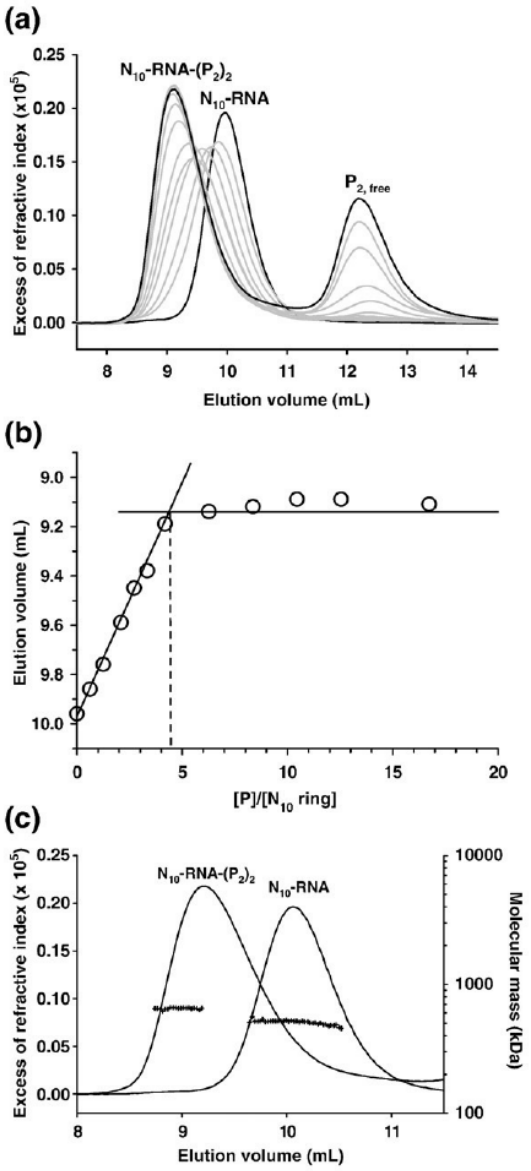


Fig. 1. Titration of N_{10}–RNA by P. (a) Chromatographic profiles at varying concentrations of full-length P. The concentration of N_{10}–RNA was kept constant at 3 μM. The samples were separated on a S200 Superdex column equilibrated in 20 mM Tris–HCl and 150 mM NaCl, pH 7.5. The black lines show the initial N10–RNA complex and the N10–RNA complex at saturation with P_2 at a [P]/[N_{10}–RNA] ratio equal to 16.7. The gray lines show the chromatograms at [P]/[N_{10}–RNA] ratios equal to 0.6, 1.3, 2.1, 2.7, 3.3, 4.2, 6.3, 8.4, 10.5, and 12.6. (b) Titration plot. The maximum volume of the peak corresponding to the N_{10}–RNA–$(P_2)_2$ complexes from the titration series was plotted as a function of the [P]/[N_{10}–RNA] ratio. (c) Molecular mass determined using SEC and online detection by MALLS and RI. The lines show the elution profile of N_{10}–RNA and of N_{10}–RNA–$(P_2)_2$ at saturation. The crosses show the molecular masses calculated from light-scattering intensity at different angles and refractive index as a function of the elution volume.

performed with the titration series recorded with N_9–RNA, N_{11}–RNA, and N_{12}–RNA. The Stokes' radii of the complexes with one or two P dimers are reported in Table 1.

Titrations monitored by SEC were also recorded with the C-terminal domain of P (P_{CTD}) and with a P mutant deleted from the central dimerization domain ($P_{\Delta 91–131}$). Both fragments of P were shown to be monomeric and to bind to N_m–RNA complexes.[9] Figure 3a shows the chromatographic series obtained with N_{10}–RNA at varying concentrations of $P_{\Delta 91–131}$. For a [$P_{\Delta 91–131}$]/[N_{10}–RNA] ratio up to 2, the peak corresponding to N_{10}–RNA–$(P_{\Delta 91–131})_n$ complexes moved to lower elution volumes with increasing concentrations of $P_{\Delta 91–131}$ (from 9.8 to 9.3 ml) and no free $P_{\Delta 91–131}$ was detectable, whereas for a [$P_{\Delta 91–131}$]/[N_{10}–RNA] ratio higher than 2, the elution volume of the N_{10}–RNA–$(P_{\Delta 91–131})_n$ complex remained constant and a peak corresponding to free $P_{\Delta 91–131}$ appeared (Fig. 3a). The titration curve (Fig. 3b) and the weight-averaged molecular mass ($\langle M_w \rangle$) measured at saturation of $P_{\Delta 91–131}$ (Fig. 3c) confirmed these observations, indicating that only two monomers bound to the N_{10}–RNA complex. Although the measured molecular masses were lower than the calculated ones (Table 1), the difference between the measured mass for N_{10}–RNA and that for N_{10}–RNA–$(P_{\Delta 91–131})_n$ at saturation by $P_{\Delta 91–131}$ was 55 kDa (565–510 = 55 kDa), corresponding closely to the difference expected for two bound $P_{\Delta 91–131}$ (calculated difference = 2 × 29.3 = 58.6 kDa). With P_{CTD}, the difference between the elution volume of N_{10}–RNA and that of N_{10}–RNA–$(P_{CTD})_n$ at saturation by P_{CTD} was too small to draw a titration curve, but the difference in molecular mass between N_{10}–RNA and N_{10}–RNA–$(P_{CTD})_n$ at saturation by P_{CTD} (532–510 = 22 kDa) measured by SEC–MALLS–RI also suggested the binding of two P_{CTD} per ring (calculated difference = 2 × 14 = 28 kDa) (Fig. 3d and Table 1).

In conclusion, these results showed that only two dimers of P or two monomeric deletion mutants containing the C-terminal domain (P_{CTD}) could bind with a high affinity to the circular N_m–RNA complexes.

Binding of P_{CTD} to circular N_{10}–RNA complex

The binding of P_{CTD} to circular N_{10}–RNA was quantified by SPR. The N_{10}–RNA complex was immobilized on a CM5 sensor chip, and P_{CTD} was injected through the flow cell. Both association and dissociation processes exhibited complex kinetic behaviors that could not be fitted with simple models and, therefore, the dissociation constant was determined by equilibrium measurements. At each protein concentration, multiple sequential injections were performed until equilibrium was reached. The binding curves could be fitted with a simple binding isotherm, yielding a dissociation constant of 160 ± 20 nM (Fig. 4a). The data were obtained in the presence of 150 mM NaCl, ruling out nonspecific binding of P_{CTD} to the chip. The

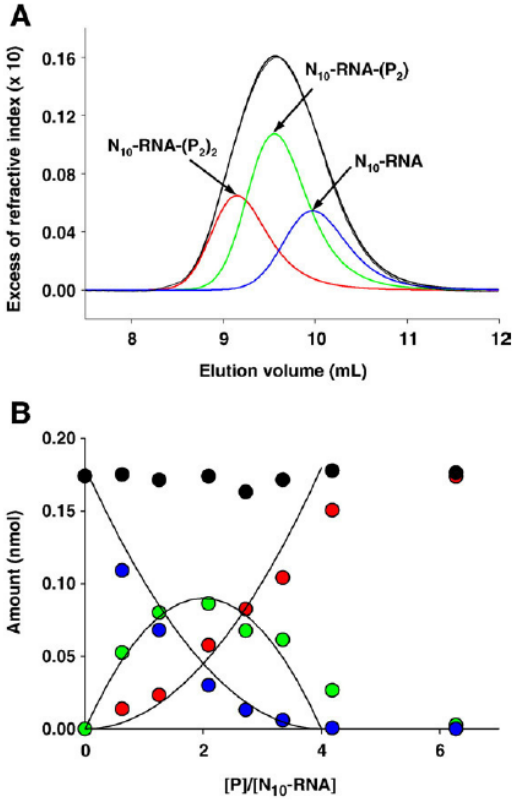


Fig. 2. Population distribution throughout the titration series. (a) Deconvolution procedure. The chromatographic peak containing the N_{10}–RNA–$(P_2)_n$ complexes was deconvoluted into three components using the routines of PeakFit according to the procedures described in Materials and Methods. The thick black line shows the experimental profile. The thin black line shows the fitted curves, and the blue, green, and red lines show the three components corresponding to $n=0$, 1, or 2, respectively. (b) Variation of the amounts of the different species as a function of the [P]/[N_{10}–RNA] ratio. The blue circles show the amount of N_{10}–RNA, the green circles the amount of N_{10}–RNA–(P_2), and the red circles the amount of N_{10}–RNA–$(P_2)_2$. The black circles show the sum of the three species. The lines show the theoretical amount estimated from a simple statistical distribution assuming two independent binding sites on N_{10}–RNA for P dimers.

Scatchard plot was linear over the used concentration range, suggesting equivalent and independent binding sites for P_{CTD} (Fig. 4b). Similar data, yielding a dissociation constant of 150±30 nM, were obtained when long helical viral NCs were immobilized on the chip and titrated with P_{CTD} (data not shown).

SAXS experiments

The RV P is a flexible protein made of structured domains separated by disordered regions.[9,10] Our attempts at crystallizing either an N_m–RNA–$(P_{CTD})_2$ or an N_m–RNA–$(P_2)_2$ complex or at soaking P_{CTD} into crystals of N_m–RNA failed. To obtain structural information about the organization of these complexes, we turned to small-angle scattering experiments and molecular modeling. SAXS profiles were recorded for N_{10}–RNA and N_{11}–RNA, for N_{10}–RNA–$(P_{CTD})_2$ and N_{11}–RNA–$(P_{CTD})_2$ at saturation by P_{CTD}, and for N_{10}–RNA–$(P_2)_2$ and N_{11}–RNA–$(P_2)_2$ at saturation by P_2 (Fig. 5a and c). For the different complexes, the scattering profiles were independent of concentration in the range of 1 to 5 mg ml^{-1}. The Guinier plots were linear for Q values ranging from 0.07 to 0.17 nm^{-1} ($Q.R_g<1.5$) (Supplementary Fig. S3) and yielded radius of gyration values (R_g) in agreement with R_g values calculated from the distance distribution functions [$P(r)$] (Table 1).

Binding of two P_{CTD} to the N_m–RNA complexes led to a slight increase of the Stokes' radius, R_S, but had no measurable effect on the radius of gyration, R_g, or on the maximum dimension, D_{max} (Table 1), thus ruling out binding of P_{CTD} to the external side of the N_m–RNA rings, which would increase R_g and D_{max}. The shape of the distance distribution function, $P(r)$, was similar for the N_m–RNA ring without or with the P_{CTD} attached, with a maximum distance probability near 9.1 nm and a shoulder near 5.0 nm (Fig. 5b and d). Conversely, addition of two P dimers led to significant increases of R_S, R_g, and D_{max} (Table 1). The maximum probability of the distance distribution function was shifted from 9.1 to 9.4 nm, and the appearance of a tail at high distance values indicated a significant enlargement of the complex (Fig. 5b and d).

Solution structure of circular N_{10}–RNA and N_{11}–RNA complexes

The structure of the N_{11}–RNA complex was solved recently by X-ray crystallography [Protein Data Bank (PDB) code: 2gtt],[7] but the R_g value calculated from the crystal structure (6.1 nm) was slightly lower than the measured R_g value (6.3±0.1), and the scattering profile calculated from the crystal structure of N_{11}–RNA showed discrepancies with the measured scattering profile (χ value=0.755) (Fig. 6a). Two flexible regions of N that were missing in the crystal structure, the N_{NTD} loop (amino acids 105–118) and the N_{CTD} loop (amino acids 376–397), were constructed in the crystallographic structure by molecular modeling in order to improve this fit (Fig. 6b). Different loop conformers were generated with the program LOBO,[18] and the best fit to the data (χ value=0.397) was obtained by adding three different loop conformers to the different protomers of N_{11}–RNA (Fig. 6a). The R_g value calculated for the best-scoring model was 6.2 nm, thus improving the agreement with the measured R_g value.

A model for the circular N_{10}–RNA complex was built by symmetrical docking with the program SYMMDOCK. An N_1–RNA subunit extracted from the crystal structure of the N_{11}–RNA complex[7] was used as the initial subunit. In the N_{11}–RNA complex, the angle between two adjacent N protomers is 147°, while in the N_{10}–RNA complex, it should be 144°.

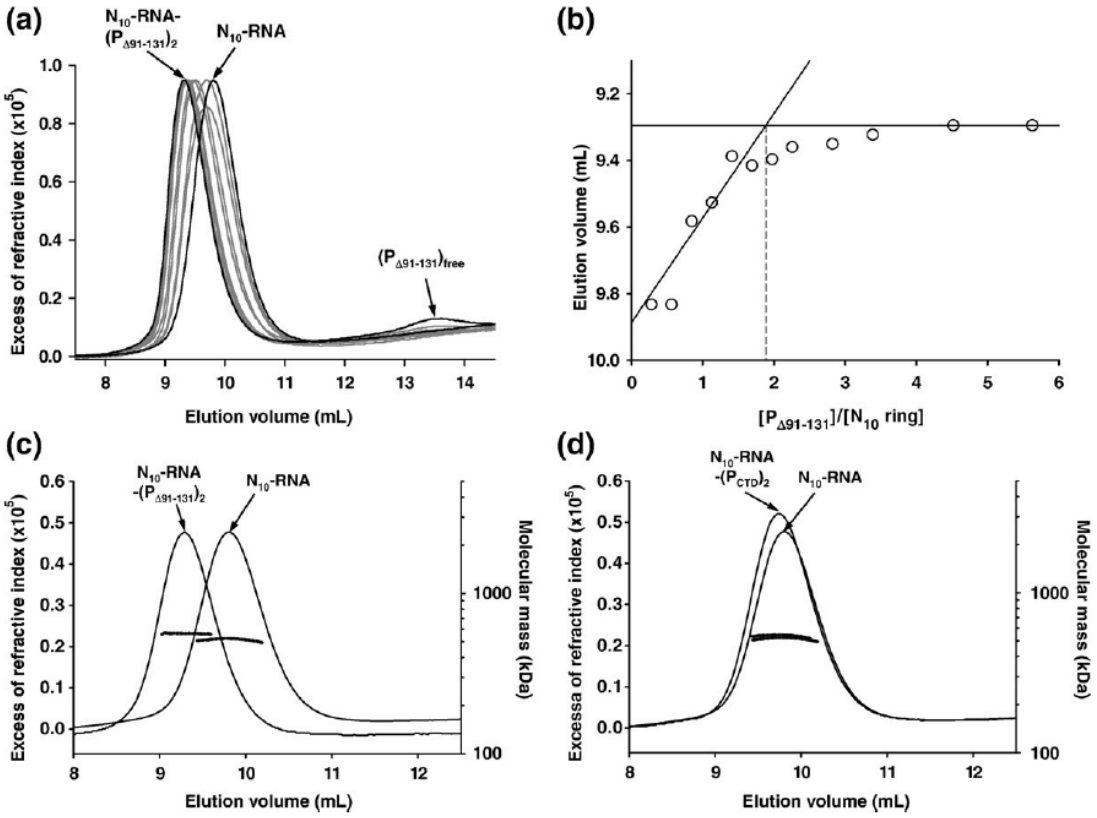


Fig. 3. Titration of N_{10}–RNA by $P_{\Delta 91-131}$ and P_{CTD}. (a) Chromatographic profiles at varying concentrations of $P_{\Delta 91-131}$. The concentration of N_{10}–RNA was kept constant at 3 μM. The samples were separated on a S200 Superdex column equilibrated in 20 mM Tris–HCl and 150 mM NaCl, pH 7.5. The black lines show the initial N_{10}–RNA complex and the N_{10}–RNA complex at saturation with $P_{\Delta 91-131}$ at a $[P_{\Delta 91-131}]/[N_{10}\text{–RNA}]$ ratio equal to 5.6. The gray lines show the chromatograms at $[P_{\Delta 91-131}]/[N_{10}\text{–RNA}]$ ratios equal to 0.3, 0.6, 0.8, 1.1, 1.4, 1.7, 2.0, 2.3, 2.8, 3.4, and 4.5. (b) Titration plot. The maximum volume of the peak corresponding to the N–RNA complexes from the titration series was plotted as a function of the $[P_{\Delta 91-131}]/[N_{10}\text{–RNA}]$ ratio. (c) Molecular mass of N_{10}–RNA–$(P_{\Delta 91-131})_n$ determined using online detection by MALLS and RI. The lines show the elution profile of N_{10}–RNA and of N_{10}–RNA–$(P_{\Delta 91-131})_2$ at saturation by $P_{\Delta 91-131}$ monitored by RI. The crosses show the molecular masses calculated from light-scattering intensity at different angles and refractive index as a function of the elution volume. (d) Molecular mass of N_{10}–RNA–$(P_{CTD})_2$ determined by SEC using online detection by MALLS and RI. The lines show the elution profile of N_{10}–RNA and of N_{10}–RNA–$(P_{CTD})_2$ at saturation monitored by RI. The crosses show the molecular masses calculated from light-scattering intensity at different angles and refractive index as a function of the elution volume.

Normal modes for N were determined using the elNémo server[19] to generate the conformational changes in the N subunit necessary to accommodate 10 subunits per ring rather than 11. A combination of two low-frequency normal modes was used to generate 20 different conformers of N in which the two sub-domains of N involved in domain exchange between adjacent protomers (amino acids 1–31 and 349–397) were moved towards the center of the ring (Fig. 6e). Models for the entire N_{10}–RNA ring were generated from the different N conformers by SYMMDOCK, and loops were added for the two flexible regions of N as for the N_{11}–RNA complex. The models were then subjected to a selection by comparing the scattering profile calculated from each model with the experimental scattering profile (Fig. 6c). The best model exhibited a χ value = 0.244 (Fig. 6d). The quality of the structure, tested with PROCHECK,[20] was similar to that of the initial crystal structure, and a comparison with the crystal structure of VSV N_{10}–RNA complex[8] yielded a C^{α} r.m.s.d. value of 0.22 nm.

Modeling of N_{10}–RNA–$(P_{CTD})_2$ and N_{11}–RNA–$(P_{CTD})_2$ complexes in solution

Titration experiments revealed that only two monomers of P_{CTD} could bind to a circular N_m–RNA complex. With this constraint, models for the N_{10}–RNA–$(P_{CTD})_2$ and N_{11}–RNA–$(P_{CTD})_2$ complexes were generated by molecular modeling using the known structure of P_{CTD}[14] and the models for the N_{10}–RNA or N_{11}–RNA complexes described above. Molecular modeling was performed in a multiple-stage approach, and the selection of models was based on SAXS and biochemical data. In a first stage, a global search for locating the binding site of P_{CTD} on the surface of the N_m–RNA

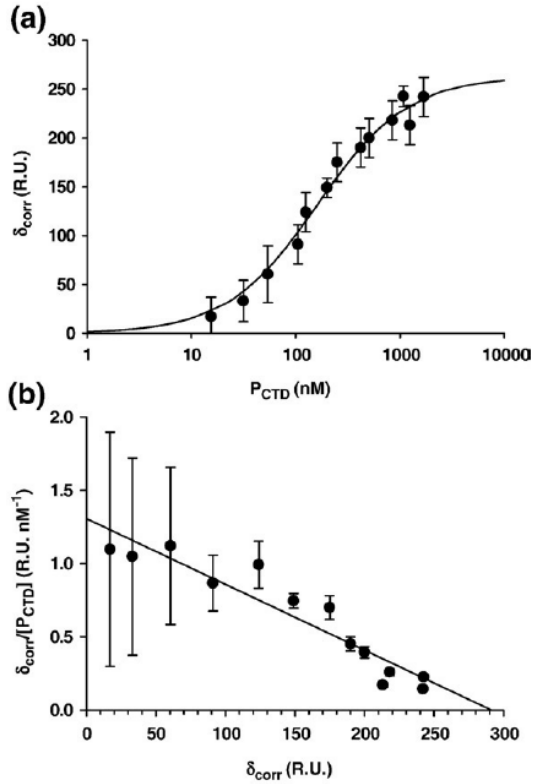


Fig. 4. Binding of P_{CTD} to N_{10}–RNA measured by SPR. (a) Equilibrium binding isotherm. Multiple injections of P_{CTD} were performed at each protein concentration until equilibrium was reached. The line was drawn using Eq. (2) and the following parameters: $(RU)_{max} = 260 \pm 10$ and $K_d = 160 \pm 20$ nM. (b) Scatchard plot. The correlation coefficient for the linear regression is equal to 0.94.

complexes was carried out by rigid-body docking with ZDOCK, a fast Fourier transform-based protein docking program[21,22] using a simplified model of the N_m–RNA complexes containing only three adjacent protomers of N–RNA–RNA, the N_3–RNA model. The generated decoys were ranked using the ZDOCK scoring function based on pairwise shape complementarity, desolvation, and electrostatics that was designed to perform best for the initial stage of unbound docking.[22] The server returned the top 2000 ranked predictions, many of which had the P_{CTD} located on the C-terminal top of the N_m–RNA protomers. Models of the entire circular complexes with two P_{CTD} bound were subsequently reconstructed by comparing the calculated and experimental scattering curves. For each decoy, different models of the entire complex were generated by positioning two P_{CTD} at different relative places on the rings (Supplementary Fig. S4). The models exhibiting the best fits to the experimental scattering curve (lowest χ values) had P_{CTD} bound to the C-terminal top of the N_m–RNA ring and the two P_{CTD} positioned opposite each other on the N_m–RNA ring. Previous observations showed that N_m–RNA rings treated with trypsin lost the C-terminal part of the N protomers (amino acids 377–450) and, consequently, their ability to bind to P_{CTD},[4,15] in good agreement with our gross localization of P_{CTD} to the C-terminal top of the ring.

The involvement of the flexible N_{CTD} loops in the binding site for P_{CTD} suggested that these loops could mold around their partner upon binding. In a second stage, to account for such an induced-fit mechanism, the structure of the complex was refined by flexible cross-docking[23] using an even more simplified model of N–RNA containing three adjacent C-terminal domains of N, the $(N_{CTD})_3$ model (Supplementary Fig. S5). First, the models were filtered by using biochemical information. Residues in P_{CTD} involved in the binding to the N–RNA complexes were previously identified through a mutational screen in the P_{CTD} of the homologous Mokola virus (63% identity between P_{CTD} of RV and Mokola virus).[24] Decoys exhibiting the largest number of contacts between equivalent residues in RV P_{CTD} and the $(N_{CTD})_3$ model were selected. Second, the selected models (65 different models) were ranked with the EMPIRE energy function,[25] and the 10 best-scoring models were refined by 40 ns molecular dynamics (MD) simulations (Supplementary Fig. S6). Third, these 10 models were used for reconstructing entire circular complexes with two P_{CTD} bound, positioning the P_{CTD} at different relative places on the rings (Supplementary Fig. S4). The decoys were then selected by comparing the calculated and measured X-ray scattering curves. For the 10 models, the best fits (lowest χ values) were obtained with the P_{CTD} bound opposite each other on the ring. In all models, the P_{CTD} were located at the C-terminal top of the N_m–RNA ring and were pinched between the N_{CTD} loops of two adjacent N protomers (Fig. 7 and Table 2). More or less, severe overlap of the two P_{CTD} domains led to higher χ values (models 4, 5, 7, 8, and 10). Figure 7 shows the three nonredundant types of arrangement of the P_{CTD} that showed no overlap (models 1, 3, and 6) together with the fitted scattering curves (Fig. 7a and e). Models where the P_{CTD} were inserted inside the ring (model 6, Fig. 7d and h) had higher χ values than those where the P_{CTD} lay on the top of the ring (models 1 and 3, Fig. 7b and c and Fig. 7f and g). Models 1 and 3, in which the P_{CTD} was lying on the top of the ring but in opposite orientations, could not be distinguished on the basis of SAXS data. In model 3 [$\chi = 0.351$ for N_{10}–RNA–$(P_{CTD})_2$ and $\chi = 0.217$ for N_{11}–RNA–$(P_{CTD})_2$], the N-terminal extremity points towards the outside of the ring (Fig. 7c and g) and incorporation of more than two P_{CTD} in the complex would be possible without steric clashes. Conversely, in model 1 [$\chi = 0.452$ for N_{10}–RNA–$(P_{CTD})_2$ and $\chi = 0.188$ for N_{11}–RNA–$(P_{CTD})_2$], the N-terminal extremity of P_{CTD} domain points towards the center of the ring (Fig. 7b and f). Incorporation of more than two P_{CTD} in this orientation led to steric clashes, in agreement with our experimental observations that only two P_{CTD} could bind to one ring. If, despite steric clashes, a third P_{CTD} was included

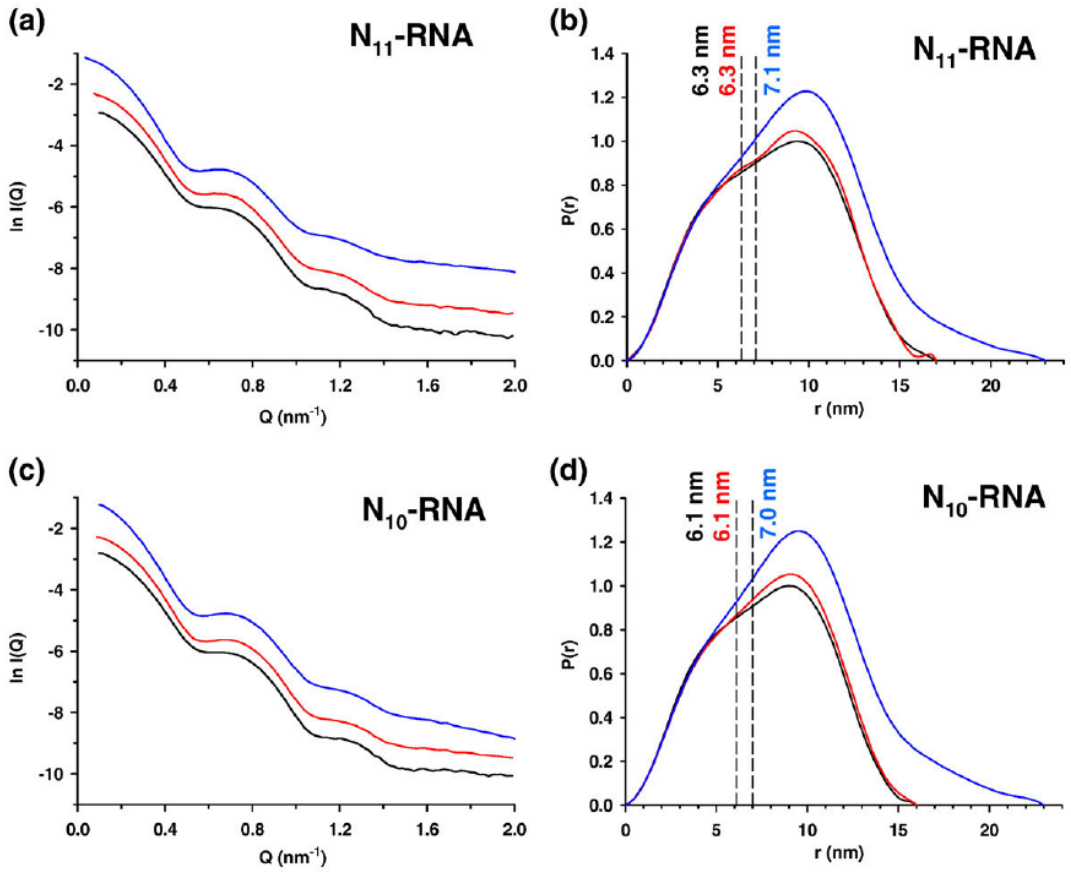


Fig. 5. SAXS experiments—scattering curve (a) and distance distribution functions $P(r)$ (b) for N_{11}–RNA complexes. The scattering profiles were recorded for N_{11}–RNA (black line), N_{11}–RNA–$(P_{CTD})_2$ (red line), and N_{11}–RNA–$(P_2)_2$ (blue line) complexes. Scattering curve (c) and distance distribution functions $P(r)$ (d) for N_{10}–RNA complexes. The scattering profiles were recorded for N_{10}–RNA (black line), N_{10}–RNA–$(P_{CTD})_2$ (red line), and N_{10}–RNA–$(P_2)_2$ (blue line) complexes. The distance distribution functions were calculated with GNOM by using D_{max} values shown in Table 1. The broken lines show the R_g value of the N_m–RNA complex and of the N_m–RNA–$(P_2)_2$ complexes.

in models 1, 3, or 6, the fitting of the calculated scattering profile to the experimental profile yielded higher χ values. Therefore, model 1, with the N-terminal extremity pointing towards the center of the ring, was selected as our best model for both N_{10}–RNA–$(P_{CTD})_2$ and N_{11}–RNA–$(P_{CTD})_2$ complexes.

In model 1, the formation of the complex occurs through extensive contacts between P_{CTD} and two adjacent N_{CTD} (buried surface, 34 nm^2). P_{CTD} is positioned above one protomer of N (N_i) (Fig. 8a), with the C-terminal helix α_6 of P_{CTD} (amino acids 279–297) packing against helix α_{14} (amino acids 402–413) (Fig. 8b). P_{CTD} is pinched between the flexible N_{CTD} loop (amino acids 376–397) of protomer N_i and of the adjacent protomer N_{i-1} (Fig. 8b). RV P_{CTD} has the shape of a half-pear, with a rounded face and a flat face.[14] The N_{CTD} loop of protomer N_i binds to the flat face of P_{CTD}, contacting the C-terminal end of helix α_4 (amino acids 265–270) and molding around a region (amino acids 241–251) encompassing the C-terminal end of helix α_2, the entire helix α_3, and the connecting loop. The N_{CTD} loop of protomer N_{i-1} binds the loop connecting helix α_1 to strand β_1 (amino acids 210–214) on the rounded face of P_{CTD}. Longer MD simulations (160 ns), performed with the $(N_{CTD})_3$ model and with model 1 of the $(N_{CTD})_3$–P_{CTD} complex, showed good stability over time for protein interactions and protein conformation. An analysis of the root-mean-square fluctuations in the $(N_{CTD})_3$ complex in the absence and presence of P_{CTD} showed that the flexible loop (amino acids 376–397) in the two N_{CTD} flanking the docked P_{CTD} (N_i and N_{i-1}) became more rigid, making contacts with the two clusters of residues in P_{CTD} that are involved in binding to N–RNA (Fig. 8c and d).[14,24] Besides these conformational changes in the N_{CTD} loops, the overall structures of the $(N_{CTD})_3$ model and of the reconstructed circular N_{10}–RNA or N_{11}–RNA complexes were almost identical with the initial structures (C^α r.m.s.d. for N_{CTD} = 0.18 nm). In the crystal structure of the N_{11}–RNA complex,[7] the N_{NTD} and N_{CTD} domains of N form "jaws" that clamp around the RNA with the closest distance between the two

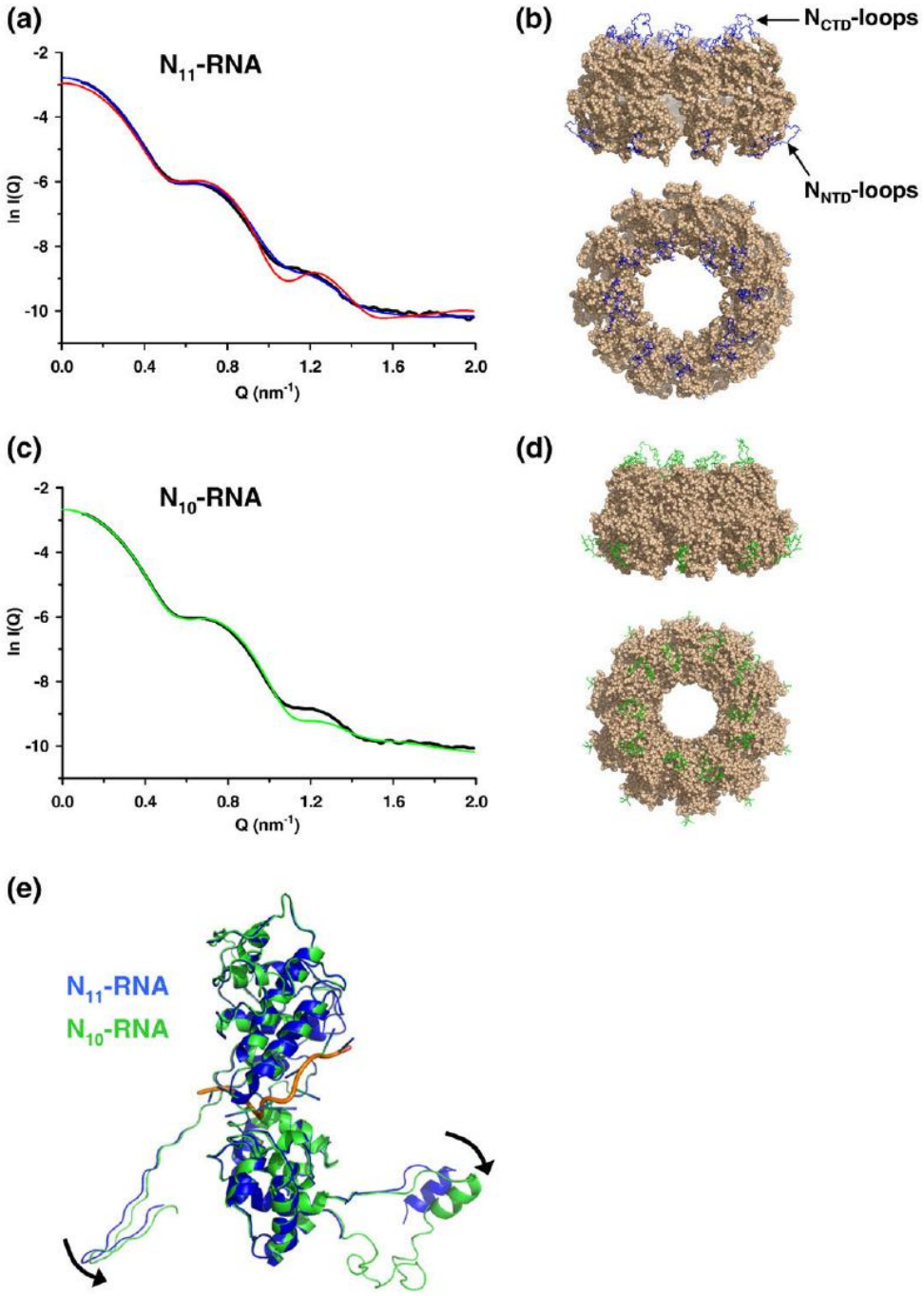


Fig. 6. Solution structure of N_{11}–RNA and N_{10}–RNA complexes. (a) SAXS profiles for the N_{11}–RNA complex. The black line shows the experimental data, the red line shows the scattering profile calculated from the crystal structure (PDB code: 2gtt), and the blue line shows the scattering profile for the best model obtained after addition of flexible loops to complete the structure. (b) Best-fit model of the N11–RNA complex. The crystal structure is shown in wheat, and the newly constructed loops are in blue. (c) SAXS profiles for the N_{10}–RNA complex. The black line shows the experimental data, and the green line shows the scattering profile for the best model obtained by adding flexible loops to complete the structure. (d) Best-fit model of the N_{10}–RNA complex. The crystal structure is shown in wheat, and the newly constructed loops are in green. (e) Ribbon diagrams of N_1–RNA complex taken from the crystal structure (blue) and of the N_1–RNA complex (green) used to construct the model for the N_{10}–RNA complex. The conformational changes were introduced in N by using normal mode motions (depicted by the arrows). The RNA fragment is shown in orange.

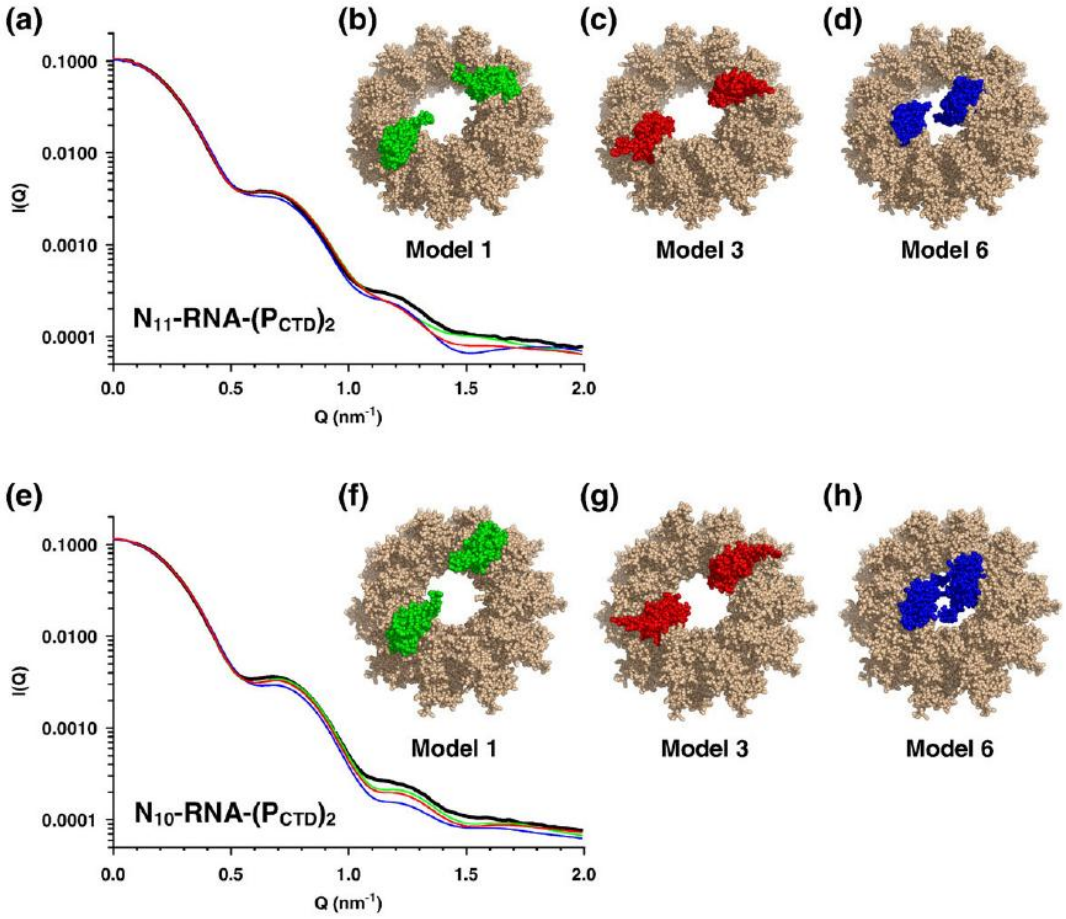


Fig. 7. Solution structure of N_{11}–RNA–$(P_{CTD})_2$ and N_{10}–RNA–$(P_{CTD})_2$ complexes. (a) SAXS profiles for N_{11}–RNA–$(P_{CTD})_2$. The black line shows the experimental data; the green, blue, and red lines show the scattering profile calculated from models 1, 3, and 6, respectively. The χ values for the fit to the experimental curves are 0.188, 0.217, and 0.708, respectively. (b–d) Location of the two P_{CTD} domains in models 1 (b), 3 (c), and 6 (d) of the N_{11}–RNA–$(P_{CTD})_2$ complex. (e) SAXS profiles for N_{10}–RNA–$(P_{CTD})_2$. The black line shows the experimental data; the blue, green, and red lines show the scattering profile calculated from models 1, 3, and 6, respectively. The χ values for the fit to the experimental curves are 0.452, 0.351, and 0.762, respectively. (f–h) Location of the two P_{CTD} domains in models 1 (f), 3 (g), and 6 (h) of the N_{10}–RNA–$(P_{CTD})_2$ complex.

jaws around 0.9 nm. To test whether this clamp was open in the complex with two P_{CTD}, we built a model for the N_{10}–RNA–$(P_{CTD})_2$ complex where the clamp was open by rotation about the hinge region, leading to an increase of the closest distance from 0.9 to 1.3 nm. This conformational change led to a large decrease in the quality of the fit to the experimental SAXS curve ($\chi = 1.23$), suggesting that binding of P_{CTD} induces no major conformational change in N (data not shown). Similarly, the structures of the bound P_{CTD} remained very similar to that of the isolated P_{CTD} (C^{α} r.m.s.d. = 0.16 nm). In P_{CTD}, the main differences (C^{α} r.m.s.d. > 0.4 nm) reside in the loop connecting strands β_1 and β_2 (amino acids 218–220), in helix α_3, and in the loop connecting helices α_2 and α_3 (amino acids 245–256), a region that is absent from the homologous C-terminal domain of VSV P.[16]

The C-terminal side of the N_m–RNA ring, in particular the N_{CTD} loop (amino acids 376–397), is rich in negatively charged residues, whereas the basic residues of the RNA binding site form a positive patch on the interior side of the ring (Fig.

Table 2. χ values obtained for the 10 best models (flexible cross-docking) by comparing the calculated and experimental scattering curves

Model	N_{10}–RNA–$(P_{CTD})_2$	N_{11}–RNA–$(P_{CTD})_2$
1	0.452	0.188
2	0.822	0.518
3	0.351	0.217
4	0.904	0.914
5	0.954	0.737
6	0.762	0.708
7	0.736	0.750
8	0.774	0.832
9	0.446	0.295
10	0.955	0.776

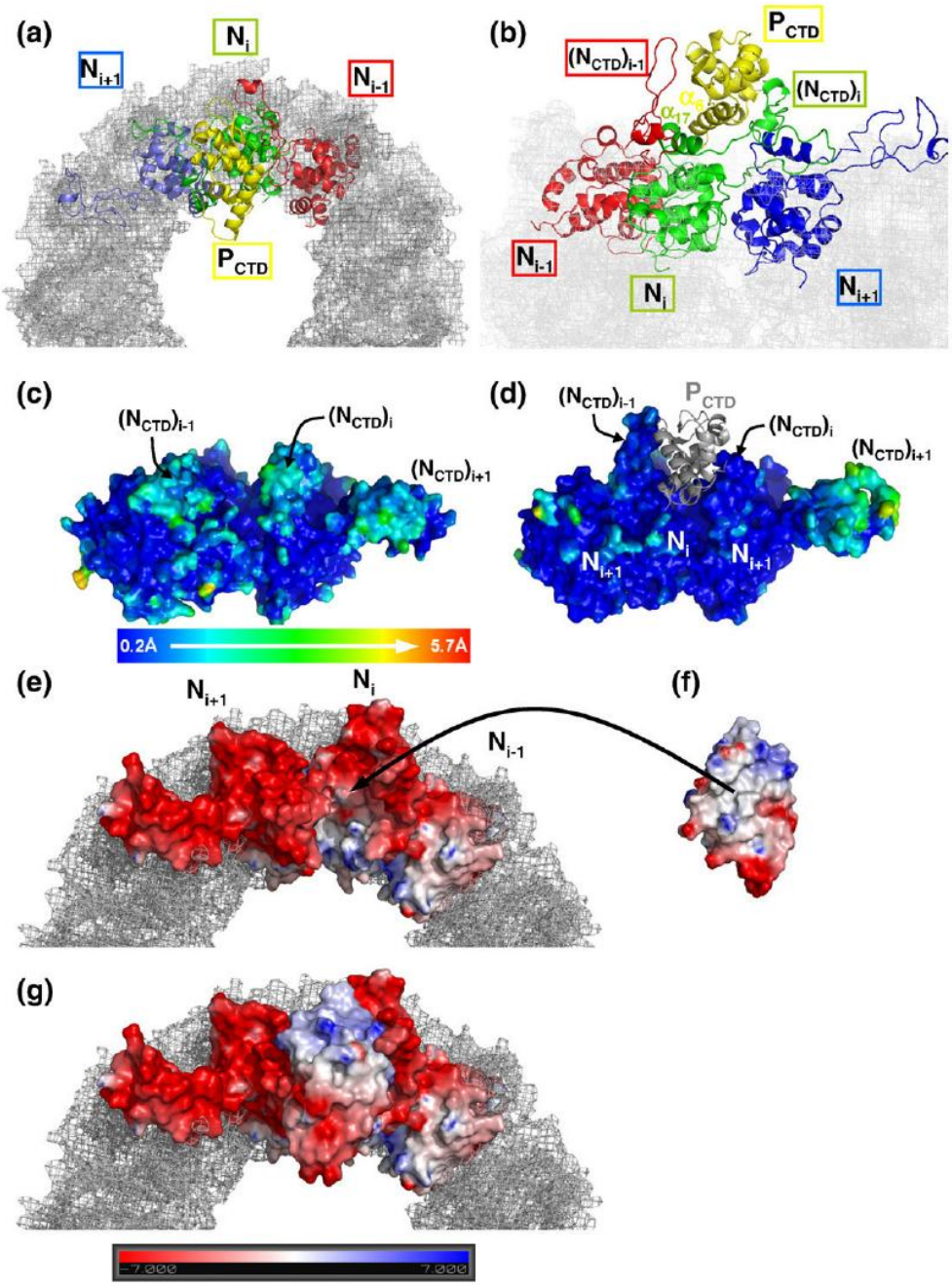


Fig. 8. Structure (model 1) of the complex formed between P_{CTD} and the N_m–RNA complex. (a) Ribbon representation showing the location of the P_{CTD} at the surface of the circular N_m–RNA ring. The black meshed structure shows the circular N_m–RNA complex. The red, green, and blue ribbon representations show the three N_{CTD} domains used in the molecular modeling, the $(N_{CTD})_3$ model. The yellow ribbon representation shows the P_{CTD}. (b) Close-up view of the interface between the P_{CTD} and the N_m–RNA surface. The N_{CTD} loop of protomers N_i and N_{i-1} are shown, as well as helix α_6 of P_{CTD} and helix α_{14} of the N_i protomer. (c and d) The root-mean-square fluctuations are mapped on the surface of the $(N_{CTD})_3$ model in the absence (c) and presence (d) of the docked P_{CTD}. The root-mean-square fluctuations were calculated from 160-ns explicit-solvent MD simulations. The root-mean-square fluctuations are mapped on the surface with blue denoting the more rigid parts and red denoting the more flexible parts. The P_{CTD} domain is shown in cartoon rendering. (e) Electrostatic surface potential of the N_m–RNA complex (including the N_{CTD} loops) (f) of the face of P_{CTD} that binds to the N_m–RNA complex and (g) of the complex between P_{CTD} and the N_m–RNA complex. The surface potentials were calculated with the Delphi program and are color-coded on the surface from red (negatively charged residues, −7 kcal/mol) to blue (positively charged residues, +7 kcal/mol). The figures were drawn with PyMOL.[26]

8e). P_{CTD} has a bipolar distribution of charges with a positive pole and a negative pole (Fig. 8f). In model 1, P_{CTD} is aligned with its positive pole lying in a cradle of negative surface area on N (Fig. 8g). Two negative patches on the surface of P_{CTD} (Asp253 and C-terminal carboxyl group) make contacts with positive patches on the interior of the ring, whereas the strongly negative patch at the N-terminus of helix α_1 of P_{CTD} (Glu191, Glu192, Asp193, and Asp194) is pointing away from the positive surface area of N, suggesting that electrostatics could serve to orient the binding of P_{CTD}.

Although molecular modeling often fails to produce high-resolution structures,[27] the models obtained for the N_m–RNA–$(P_{CTD})_2$ complexes exhibit numerous molecular interactions that could explain the high stability of the complexes and are in agreement with available biochemical data. Mapping of mutations in Mokola virus P_{CTD} that affect binding to N–RNA complexes on the RV P_{CTD} three-dimensional structure highlighted the importance of the C-terminal helix α_6 and revealed the presence of two clusters of residues on opposite faces of P_{CTD} that are critical for the interaction.[14,24] The first cluster (cluster 1) consists of three Lys residues (Lys211, Lys212, and Lys214) and Leu224. The second cluster (cluster 2) consists of Cys261, Trp265, and Met287, which forms with Leu244, Pro245, and Leu291, a hydrophobic pocket, the W-hole. In our model, the packing of the C-terminal helix α_6 of P_{CTD} (amino acids 279–297) against helix α_{14} (amino acids 402–413) involves the formation of two salt bridges between Glu403 and Arg408 of N_i and between Lys285 and Asp289 of P_{CTD}, respectively (Fig. 9a). Residues 376 to 397 from the N_{CTD} loop of the N_i protomer pack against the flat face of P_{CTD}. In particular, Val379 and Phe395 are lodged into the W-hole of P_{CTD} (cluster 2) (Fig. 9b). The N_{CTD} loop from the adjacent protomer N_{i-1} (amino acids 376–396) binds to the rounded face of P_{CTD},

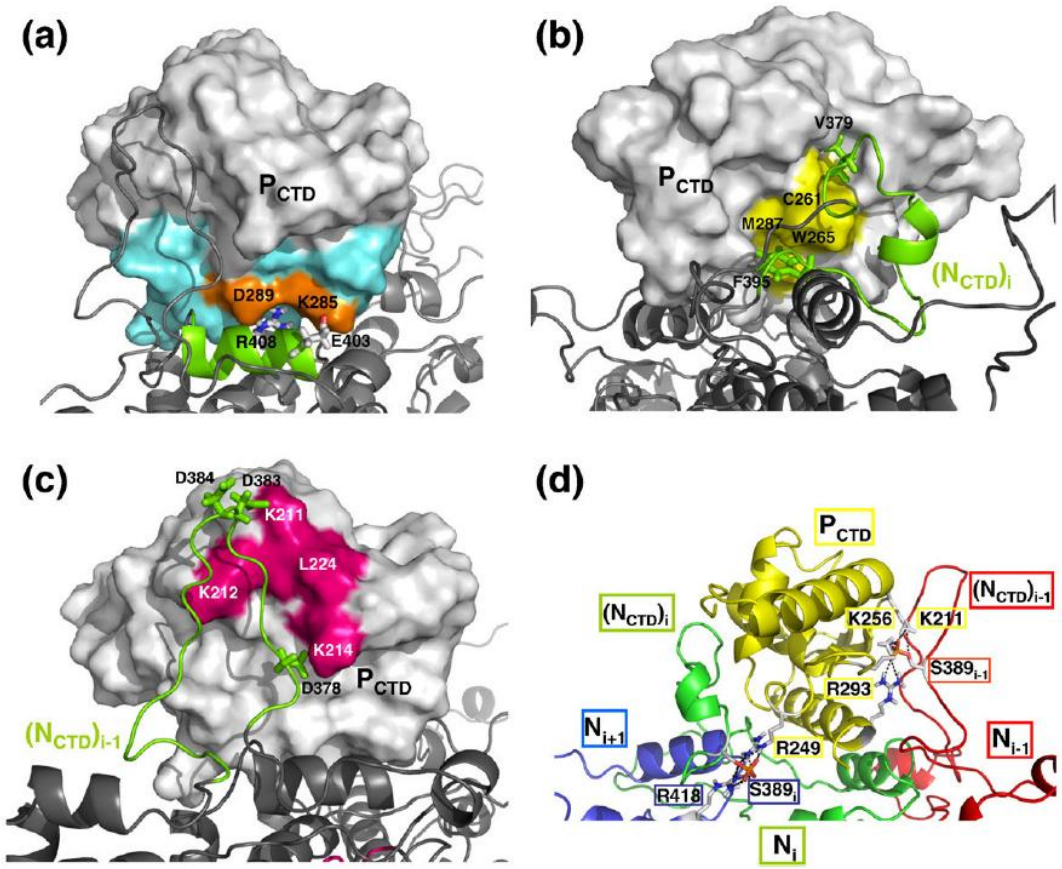


Fig. 9. Detailed interactions between N_{CTD} and P_{CTD}. (a) Helix α_6 of P_{CTD} lies on helix α_{14} of N_{CTD}. Helix α_6 of P_{CTD} is shown in cyan with residues Lys285 and Asp289 highlighted in orange. Helix α_{14} of the N_i protomer is shown in green with residues Glu403 and Arg408 shown as sticks. (b) Interactions with residues in cluster 1 of P_{CTD}. The NCTD loop of protomer N_i is shown in green. Residues Cys261, Met287, and Trp265 of P_{CTD} are shown in yellow, whereas residues Val379 and Phe395 are shown as sticks. (c) Interactions with residues in cluster 2 of P_{CTD}. The NCTD loop of protomer N_{i-1} is shown in green. Residues Lys211, Lys212, Lys214, and Leu224 of P_{CTD} are shown in pink, whereas residues Asp378, Asp383, and Asp384 of protomer N_{i-1} are shown as sticks. (d) Phosphorylation of Ser389 of N_i and N_{i-1} protomers. The phosphorylated Ser389 of the two protomers and the residues forming the two networks of ionic bonds are shown as sticks. The ionic bonds are shown as dotted lines.

with Asp378, Asp383, and Asp384 of N making ionic bonds with Lys211, Lys212, and Lys214 (cluster 1) of P_{CTD}, respectively (Fig. 9c). Phosphorylation of Ser389 in the N_{CTD} loop enhances the association with P.[28] To test the effect of this phosphorylation in our model, phosphoryl groups (Pi) were added to Ser389 of both N_i and N_{i-1} protomers in model 1 of the $(N_{CTD})_3$ complex, and the structure was submitted to a 20-ns MD simulation. During the simulation, the Pi group on the N_i and N_{i-1} protomers made networks of salt bridges with positively charged residues in P_{CTD} and N_{i+1}. The Pi group on the N_i protomer made a network of salt bridges with the side chains of Arg249 of P_{CTD} and of Arg418 of N_{i+1}, while the equivalent Pi group on the N_{i-1} protomer made a network of salt bridges with the side chains of Lys211, Lys256, and Arg293 of P_{CTD} (Fig. 9d). The formation of these interactions could explain the strengthening of the interaction. P_{CTD} also contains two phosphorylation sites for the cellular protein kinase C, Ser210 and Ser271, although the functional role of these modifications is unknown. Ser210 is part of the P_{CTD} region that binds to the N_{CTD} loop of the N_{i-1} protomer. Its phosphorylation would introduce additional negative charges in a highly charged region of the complex (Lys211, Lys212, and Lys214 of P_{CTD} and Asp378, Asp383, and Asp384 of N_{CTD}), which could either weaken or strengthen the interaction. Finally, a monoclonal antibody that recognizes P bound to the N–RNA complex with more affinity than soluble P was described.[29] The presumed epitope located around Gln275 on helix α_5 of P_{CTD} is exposed at the surface of the N_m–RNA–$(P_{CTD})_2$ complexes and is accessible to the antibody. The increased efficiency of antibody binding could result from the involvement of residues of the N_{CTD} loop in the recognized epitope.

Discussion

Binding of P_{CTD} to the N_m–RNA template in *Rhabdoviridae*

The binding site for P_{CTD} on the RV NCs was previously mapped to the C-terminal region of N (amino acids 377–450) by limited proteolysis.[4,15] Here, we applied molecular docking to build models of the complex formed between P_{CTD} and circular N_m–RNA containing 10 or 11 protomers of N and used SAXS and biochemical data to select the best models. In our models, the binding site of P_{CTD} on N is located at the C-terminal top of one N protomer (N_i) but also involves interactions with one adjacent N protomer (N_{i-1}). The C-terminal helix α_6 of P_{CTD} lies along helix α_{14} of the C-terminal domain of one N protomer (N_i), while the N_{CTD} loops from this protomer (N_i) and from the adjacent one (N_{i-1}) pinch the P_{CTD} domain. This model is in agreement with available biochemical data.

The structure of the complex explains the binding specificity of P_{CTD} for the N_m–RNA complexes. P_{CTD} cannot interact with N^{011} because its binding site is incomplete, and the closure of the adjacent N_{CTD} loops appears, thus, as an essential component of the binding process. The recent crystal structures of circular N–RNA complexes from RV and VSV revealed that the RNA bases were completely sequestered within the protein, clamped between N_{NTD} and N_{CTD} domains,[7,30] and were thus protected from chemical modifications.[31] Therefore, the genomic RNA molecule is not directly accessible to L, and it was hypothesized that binding of P to the N–RNA template could trigger the conformational change in N to release the RNA molecule and make it accessible for the RNA polymerase.[7,32] Our modeling approach revealed no evidence for major conformational changes in N or P_{CTD} upon formation of the complex. The models generated by assuming a closed conformation for the N protomers, as in the crystal structure of the N_{11}–RNA complex,[7] fitted the experimental SAXS curve. Only limited conformational changes were observed in two regions of P_{CTD} and in the N_{CTD} loop of N. However, there is no evidence that the circular N_m–RNA complexes are active templates for RNA synthesis, and it is possible that ring closure blocks the conformational change. The natural N_m–RNA template consists of flexible helical NCs, and binding of P to that structure could trigger additional conformational changes. Alternatively, the conformational change required for giving access for the RNA to the polymerase could be induced by the L subunit of the polymerase itself.

After our paper was submitted, Green and Luo reported the crystal structure of an equivalent complex from the closely related VSV.[33] The structural arrangement of P_{CTD} in the VSV complex[33] appears to be very similar to that obtained for the RV complex by molecular modeling (this study). In both VSV and RV complexes, the P_{CTD} lies on the top of an N protomer (N_i) and is pinched by N_{CTD} loops from two adjacent N monomers. The P_{CTD} exhibits the same orientation on the top side of the N–RNA ring, its N-terminal extremity pointing towards the center of the ring. In the VSV complex, the C-terminal helix α_4 of P_{CTD} occupies a similar position than the C-terminal helix α_6 of RV PCTD and interacts with helix α_{13} of VSV N that is equivalent to helix α_{14} of RV N.[7,8,32] The main differences between the RV and VSV complexes reside in the numbers of residues involved in the interaction and in the surface area buried in the complex. In the RV complex, 34 residues of P_{CTD} and 39 residues of N (28 in N_i and 11 in N_{i+1}) participate in the binding against 18 and 17 residues in the VSV P_{CTD} and N, respectively. The N_{CTD} loop is longer in RV than in VSV, and the loops from two adjacent N protomers form a larger clamp in RV that hides a larger part of P_{CTD} than in VSV. In the RV complex, the total surface area of P_{CTD} buried upon interaction with two adjacent N molecules is 1739 Å^2 [~27% of the total surface area (6465 Å^2)], significantly larger than that in the VSV complex where only 956 Å^2 are

buried (~19% of the total surface area). Another important difference is in the occupancy of the potential P_{CTD} binding sites on the N–RNA ring. In the VSV crystal structure, one P_{CTD} is bound to each N protomer, although reduced occupancy was observed in some positions.[33] In the RV complex (this study), the titration experiments revealed two independent binding sites for isolated P_{CTD} on the circular N_m–RNA complexes containing 9 to 12 N protomers. In the structural modeling procedure, it was not possible to insert more than two P_{CTD} on the N_{10}–RNA ring without generating steric clashes, a feature that explains our experimental results with P_{CTD}, $P_{\Delta 91–131}$, and full-length P. Thus, in the RV system, the stoichiometry of the complex formed with circular N_m–RNA complexes is certainly imposed by steric hindrance and might not be indicative of the stoichiometry with natural NCs. The viral NCs form helical structure with ~15 N protomers per turn in isolated viral NCs and ~53 N protomers per helical turn in the viral particles,[4,5,32] which have a wider angle between N protomers and could possibly accommodate more P_{CTD}. In VSV, the protein composition of the virion determined by scanning transmission electron microscopy indicated 1 P dimer (233 P_2 molecules) for 5.4 N molecules (1258 N molecules),[34] a value in good agreement with that obtained with the circular N–RNA complexes. However, given the structure of the complex between RV P_{CTD} and the 2 adjacent N protomers forming its binding site, it seems unlikely to reach a stoichiometry larger than 1 P_{CTD} for 2 N protomers.

Comparison with *Paramyxoviridae*

The mechanisms of RNA synthesis by the transcription/replication complex have been extensively studied for Sendai virus, a virus belonging to another *MNV* family (*Paramyxoviridae*).[2] In the *Paramyxoviridae*, P also acts as a cofactor of the polymerase and mediates the attachment of the L protein to the N–RNA template. Sendai virus P forms tetramers through its central coiled-coil domain and attaches to the N–RNA template through its C-terminal PX domain.[35] PX is a three-α-helix bundle that binds to the C-terminal disordered region of N, named N_{TAIL}, in a region that populates α-helical conformers[36] and presumably is stabilized upon binding to PX. The current model proposes that P tetramers cartwheel along the N–RNA template, carrying along their L cargo.[2,37] For this, P would rotate around its central coiled-coil axle with the four C-terminal PX domains continuously associating and dissociating from the N–RNA template.[2,37] In agreement with such model, the affinity of PX for N_{TAIL} is weak; the dissociation constant measured for the PX/N_{TAIL} complex is around 60 μM and seems to be independent of the length of the N_{TAIL} fragment.[35,38] Similar measurements with the PX domain of measles virus yielded a dissociation constant of the same order of magnitude.[39]

A model for the attachment of L to the N_m–RNA template

With RV, a different picture is emerging. We report here that the affinity of P_{CTD} for the N_m–RNA template is high, with dissociation constants in the nanomolar range. The SEC titration experiments indicate that the dissociation constant for P_2 is also lower than micromolar levels. Moreover, the association and dissociation kinetics measured by SPR are slow (tens of seconds time range) as compared to the time scale of viral replication. Both observations are difficult to conciliate with a mechanism involving continuous association and dissociation of P_{CTD} from the N_m–RNA template. Moreover, Jacob *et al.* showed previously that the P protein mutant deleted from the central dimerization domain remained functional in transcription.[24] Recently, we showed that $P_{\Delta 91–131}$ was monomeric[9] and, thus, that dimerization is not required for RNA synthesis in contrast with the situation in the *Paramyxoviridae*.

Because of the high affinity of P_{CTD} for the N_m–RNA templates, P_2 molecules might be continuously attached to the N_m–RNA template in the virion and in the cytoplasm of infected cells. The estimated number of P_2 molecules in the virion[34] would result in one P_2 molecule attached every five N protomers. The L protein could simply move along the template by jumping between adjacent P_2 molecules.

Materials and Methods

Sample preparation

Recombinant N–RNA rings, full-length P, $P_{\Delta 91–131}$, and the C-terminal domain of P (P_{CTD}) were produced and purified as previously described.[6,10,16] RV NCs were produced and purified from infected BSR cells as described previously.[31] P protein concentrations were measured by absorbance spectroscopy using Edelhoch's method.[40] N–RNA concentrations were measured by integrating the elution peak of the SEC profile monitored by RI and using a dn/dc value of 0.185 ml g^{-1}. Samples of N_m–RNA–$(P_2)_n$ complexes were obtained by mixing P with N_m–RNA at different [P]/[N_m–RNA] ratios and incubating the mixture for 10 min at room temperature. For SAXS and SANS, samples of saturated N_m–RNA–$(P_2)_2$ complexes were obtained by mixing P with N_m–RNA in a 4:1 ratio and incubating the mixture for 10 min at room temperature. Each sample used in SAXS or SANS experiment was tested by SEC–MALLS–RI for the homogeneity of the complex by verifying the absence of peak corresponding to free P or to free N–RNA complex.

Surface plasmon resonance

Analyses were carried out on a Biacore X (GE Healthcare). N_{10}–RNA complex was immobilized on a CM5 sensor chip using an amine coupling kit (GE Healthcare) to final resonance values of about 4000 response units (RUs). P_{CTD} was in 20 mM Tris–HCl buffer at pH 7.5 containing 150 mM NaCl, 1 mM DTT, and 0.005% (v/v) Tween 20. Data were collected at 20 °C at a flow rate of 20 μl/min.

Binding curves were corrected for background and bulk refractive index contribution by subtraction of the corresponding reference flow cell. Regeneration of the surface was achieved by injection of 20 µl of 20 mM Tris–HCl buffer at pH 7.5 containing 1 M NaCl at a flow rate of 50 µl/min. Over the course of the experiment, the amount of immobilized N_{10}–RNA decreased by 10–20% between the first and the last P_{CTD} injections. The binding at each P_{CTD} (δ_{mes}) concentration was adjusted (δ_{corr}) for the level of N_{10}–RNA on the surface immediately preceding that injection using the equation:

$$\delta_{corr} = \delta_{mes} \frac{(RU_{N10-RNA})_{initial}}{(RU_{N10-RNA})_{injection}} \quad (1)$$

where $(RU_{N10-RNA})_{initial}$ is the level of immobilized N_{10}–RNA immediately preceding the first injection of P_2. The dissociation constant, K_d, was determined by nonlinear regression from equilibrium resonance measurements for P concentrations ranging from 10 to 1700 nM using the simple Langmuir binding isotherm:

$$\delta_{corr} = \frac{(RU)_{max}[P_2]}{K_d + [P_2]} \quad (2)$$

where δ_{corr} is the adjusted signal measured in RUs and $(RU)_{max}$ is the maximum response.

SEC–MALLS–RI

SEC was performed with an S200 Superdex column (GE Healthcare) equilibrated with 20 mM Tris–HCl and 150 mM NaCl, pH 7.5. Separations were performed at 20 °C with a flow rate of 0.5 ml min^{-1}. Typically, 50 µl of a protein solution at a concentration of 1–5 mg ml^{-1} was injected. Online MALLS detection was performed with a DAWN-EOS detector (Wyatt Technology Corp., Santa Barbara, CA) using a laser emitting at 690 nm. Data were analyzed, and weight-averaged molecular masses (M_w) were calculated using the ASTRA software (Wyatt Technology Corp.) as described previously.[10]

The column was calibrated with proteins of known Stokes' radii (R_S) and molecular masses (M_w)[41]: bovine serum albumin (R_S=3.4 nm, M_w=67.0 kDa), RNase A (R_S=1.9 nm, M_w=13.7 kDa), ovalbumin (R_S=3.0 nm, M_w=43.5 kDa), β-lactoglobulin (R_S=2.7 nm, M_w=36.8 kDa), and chymotrypsinogen (R_S=2.3 nm, M_w=25 kDa).[41]

The chromatographic peaks recorded by RI were modeled with PeakFit v4.12 (SeaSolve Software Inc.) using the EMG model. The EMG model, a widely used empirical chromatographic model capable of modeling tailing, is a mathematical convolution of a Gaussian function with an exponential response function that involves four parameters according to the following equation:

$$y = \frac{A}{2d}\exp\left(\frac{w^2}{2d^2} + \frac{C - V_{el}}{d}\right)\left[\mathrm{erf}\left(\frac{V_{el} - C}{\sqrt{2}w} - \frac{w}{\sqrt{2}d}\right) + \frac{d}{|d|}\right] \quad (3)$$

where A is the area under the curve, C is the elution volume corresponding to the center of the peak, w is the width at half-height, d is an empirical parameter accounting for the distortion of the peak, and erf is the Gauss error function of sigmoidal shape given by:

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt \quad (4)$$

The chromatographic titration series was deconvoluted into four components in three steps. In a first step, the elution peaks corresponding to the two components that could be prepared in a homogenous form, N_{10}–RNA alone and N_{10}–RNA–$(P_2)_2$ at saturation of P, were modeled with a single EMG function. In a second step, three parameters (C, w, and d) of each of these three components were kept constant and the automatic deconvolution process of PeakFit was used to determine the parameters corresponding to the fourth component, N_{10}–RNA–(P_2), from chromatograms obtained at intermediate concentrations of P. In the third step, the three parameters, C, w, and d, for the four components were fixed and all chromatograms of the titration series were fitted to determine the area (A) for each component as a function of the [P]/[N_{10}–RNA] ratio. The total surface area varied linearly with the total amount of N_{10}–RNA and P with a correlation coefficient of 0.992 (Supplementary Fig. S2a). Assuming an identical dn/dc value for the different components (0.185 ml g^{-1}), the molecular masses of N_{10}–RNA and P were used for converting the surface area of the different peaks into amounts in nanomoles. The graph of the calculated amount of P, including free or bound forms [2× amount of N10–RNA–(P_2)+4× amount of N_{10}–RNA–$(P_2)_2$], increases linearly with the amount of added P with a slope of 1.00 and a correlation coefficient of 0.994, whereas the graph of the calculated amount of N_{10}–RNA, including free and bound forms [amount of N_{10}–RNA–(P_2)+amount of N_{10}–RNA–$(P_2)_2$], remained constant for the different [P]/[N_{10}–RNA] ratios (Supplementary Fig. S2b).

Assuming that the titration was performed at concentrations of P higher than K_d, every added P molecule bound to N–RNA until saturation was reached. For two binding sites per ring, the statistical distribution of the different forms [N_{10}–RNA, N_{10}–RNA–(P_2), and N_{10}–RNA–$(P_2)_2$] was obtained by multiplying adequately the relative probabilities of finding a bound P_2 or an empty site at each site of the N–RNA complex. For example, if we have 100 molecules of N–RNA (200 binding sites) and 40 molecules of P_2, the probability of finding one P_2 bound to one of the two N–RNA binding sites is P (bound)=40/200. Therefore, the probability of finding an empty site is P (empty)=1−P (bound). The fractions of the different forms of the complex are then given as follows:

$$f_{N10-RNA} = \mathcal{P}(\text{empty}) \times \mathcal{P}(\text{empty}) = 0.8 \times 0.8 = 0.64 \quad (5a)$$

$$f_{N10-RNA-(P2)} = [\mathcal{P}(\text{empty}) \times \mathcal{P}(\text{bound})] \times 2 = [0.8 \times 0.2] \times 2 = 0.32 \quad (5b)$$

$$f_{N10-RNA-(P2)2} = \mathcal{P}(\text{bound}) \times \mathcal{P}(\text{bound}) = 0.2 \times 0.2 = 0.04 \quad (5c)$$

Small-angle X-ray scattering

SAXS data were collected at the European Synchrotron Radiation Facility (Grenoble, France) on the beamlines ID2[42] and ID14-3. The sample-to-detector distances were 1 and 5 m on ID2 and 1 m on ID14-3. The wavelength of the X-rays was 0.0995 nm so as to cover the angle region characterized by Q values ranging from 0.10 to 4.70 nm^{-1} on ID2 and from 0.07 to 3.5 nm^{-1} on ID14-3. Samples were contained in a 1.9-mm-wide quartz capillary. The time of exposure was optimized for reducing radiation damage.

Protein concentrations were between 1 and 5 mg ml^{-1}. Reduction of the 2D SAXS patterns to 1D scattering profiles was performed using the established procedure available at ID2 and ID14-3, and buffer background runs were subtracted from sample runs.

Radius of gyration and forward intensity at zero angle [$I(0)$] were determined with the program PRIMUS[43] by using Guinier approximation at low Q values, corresponding to a range of $Q.R_g$ values up to 1.5:

$$\ln I(Q) = \ln I(0) - \frac{R_g^2 Q^2}{3} \quad (6)$$

The radius of gyration and the pairwise distance distribution function $P(r)$ were calculated by indirect Fourier transform with the program GNOM.[44] The maximum dimension (D_{max}) value was adjusted so that the R_g value obtained from GNOM agrees with that obtained from the Guinier analysis.

Structure modeling

The LOBO,[18] SYMMDOCK,[45] elNémo,[19] ZDOCK,[21,22] and EMPIRE[25] servers and the BiGGER software[46] were used with their default settings. The calculations of the scattering curves from the different models and the comparison with the experimental scattering curves were performed with CRYSOL.[47]

Models for the structure of the N_{10}–RNA–$(P_{CTD})_2$ and N_{11}–RNA–$(P_{CTD})_2$ complexes were generated by docking the P_{CTD} domain (PDB code: 1vyi) to the N_{10}–RNA or N_{11}–RNA complexes derived from the crystal structure of N_{11}–RNA (PDB code: 2gtt). A multiple-stage docking procedure was used, and the models were selected on the basis of available biochemical data and on a comparison of the calculated and experimental X-ray scattering profiles.

Rigid-body docking

In a first stage, the gross location of the P_{CTD} domains on the N_m–RNA complex was determined by a regular rigid-body docking procedure. Because of its small size, the binding site for one P_{CTD} cannot involve more than three adjacent protomers of N, and thus, to optimize computational time, we worked with a simplified model containing only three adjacent N protomers, named the N_3–RNA model. The docking of P_{CTD} on this trimeric N_3–RNA complex with the ZDOCK server resulted in an unbiased ensemble of conformers from which the 100 best-scored decoys were selected by using the scoring function of the ZDOCK server.[21,22] Circular N_{10}–RNA–$(P_{CTD})_2$ or N_{10}–RNA–$(P_{CTD})_2$ complexes were generated by replacing two N_3–RNA trimers from the N_m–RNA model with two N_3–RNA–P_{CTD} complexes. The N_3–RNA–P_{CTD} trimers were introduced at different positions in the ring (Supplementary Fig. S4), in order to sample all possible combinations of relative positioning of the two P_{CTD}. These models were subjected to a selection by comparing the calculated and experimental scattering profiles using the program CRYSOL.[47]

Flexible docking

In a second stage, an implicitly flexible cross-docking approach was used to refine the predicted structure of the complexes.[48] Supplementary Fig. S5a shows a scheme of the docking protocol. Again, the trimeric N_3–RNA model was further simplified by removing the N-terminal domain of N (amino acids 1–234 and 274–300) and the RNA fragment to optimize computational time. The initial $(N_{CTD})_3$ model and the crystal structure of P_{CTD} (PDB code: 1vyi) were submitted to 10-ns MD simulations with explicit solvent using GROMACS 4.0.2 software package[49] and the GROMOS 53a6 force field.[50,51] To avoid unwanted motions of the free chain extremities resulting from the truncation of the proteins, we imposed constraints on the C^α of residues 349 to 360, which protrude from each protomer and make tight contacts with the adjacent protomer. At the beginning of each simulation, the protein was immersed in a box of SPC/E water molecules. Sodium and chloride ions were added to reach a 150-mM salt concentration. Long-range electrostatics was calculated with particle-mesh Ewald summation. Hydrogen atoms were treated as virtual sites, enabling a 5-fs integration time step to be used. The v-rescale thermostat[52] and the Parrinello–Rahman barostat[53] were used to maintain a temperature of 300 K and a pressure of 1 atm. The system was equilibrated for 125 ps with restrained protein atoms before the beginning of the simulation. Resulting 10-ns-long trajectories for $(N_{CTD})_3$ and P_{CTD} were clustered according to their r.m.s.d. using g_cluster,[54] and 10 snapshots were extracted and used as input for rigid-body docking. Cross-docking calculations (10×15) between one representative of each cluster was performed with BiGGER[46] using the default parameters. A map showing the location of P_{CTD} on the $(N_{CTD})_3$ complex in the first 1000 best solutions generated with BiGGER for a typical cross-docking run showed that in the best-scoring models (red spheres), the docked P_{CTD} was located between the flexible loops of two adjacent N_{CTD} domains (Supplementary Fig. S5b).

After each run, the 1000 solutions were ranked using available biochemical data and the global score of BiGGER. Two clusters of residues on two opposite faces of P_{CTD} had been previously involved in binding to N_m–RNA by mutagenesis in a closely related virus.[14,24] A contact score was used to select models that contain a maximum of N to P contacts involving these residues of P_{CTD} (K239, K256, K272, K282, K211, K212, K214, L224, C261, W265, and M287) and any residue in the $(N_{CTD})_3$ complex. The best-scoring model for each of the 150 individual runs was clustered in order to remove similar decoys, yielding an ensemble of 65 decoys that were then subjected to steepest descent and conjugate gradient energy minimization with the GROMACS 4.0.2 software package.[49]

The minimized structures were ranked with the EMPIRE interaction energy function,[25] and the 10 best-scoring models were selected for further MD refinement. These models were subjected to 10 ns (models ranked 6 to 10) or 40 ns (models ranked 1 to 5) of explicit-solvent MD simulations using the procedure described above. Again, the C^α atoms of residues 349–360 of $(N_{CTD})_3$ were position restrained to maintain a shape close to that of the crystal structure. The conformational drift of the structure throughout the MD simulations measured as the C^α r.m.s.d. with respect to the initial structure is shown in Supplementary Fig. S6 for the 5 best models. After an equilibration of about 1 ns, the different models displayed stable trajectories for both the $(N_{CTD})_3$ and the P_{CTD} parts.

The 10 models refined for 10 ns by MD simulation were used to construct models for the N_{10}–RNA–$(P_{CTD})_2$ and N_{11}–RNA–$(P_{CTD})_2$ complexes. Models for the N_{10}–RNA–$(P_{CTD})_2$ and N_{11}–RNA–$(P_{CTD})_2$ complexes were constructed by replacing two $(N_{CTD})_3$ from the structure of the corresponding N_m–RNA circular complexes by one of the 10-ns-refined structural decoys of the $(N_{CTD})_3$–P_{CTD}

complex taken from the MD simulations. As in the preliminary rigid-body docking procedure, the two trimers were introduced at different positions in the ring (Supplementary Fig. S4). The scattering profile was calculated and fitted to the experimental SAXS data using CRYSOL,[47] yielding χ values ranging from 0.351 (model 3) to 0.955 (model 10) for the N_{10}–RNA–$(P_{CTD})_2$ complex and from 0.188 (model 1) to 0.914 (model 4) for the N_{11}–RNA–$(P_{CTD})_2$ complex. The best χ values were obtained when the two P_{CTD} were opposite each other on the N_m–RNA ring, with three N protomers on each side of the two protomers bound to P_{CTD} for the N_{10}–RNA–$(P_{CTD})_2$ complex (type 3 in Supplementary Fig. S4a) or with three and four N protomers on each side of the two protomers bound to P_{CTD} for the N_{11}–RNA–$(P_{CTD})_2$ complex (types 3 and 4 in Supplementary Fig. S4b).

Phosphorylated model

A model was constructed by adding phosphate groups to Ser389 of N_i and N_{i-1} protomers, and a new 50-ns MD simulation was run as described above. The topology for the phosphorylated serine residues (Sep) was generated using the PRODRG server.[55]

Acknowledgements

This work was supported by the interdisciplinary program "Maladies Infectieuses Emergentes" from the Centre National de la Recherche Scientifique and by a grant from the French ANR [ANR-07-001-01 (ANRAGE)] and Lyonbiopôle. The authors thank Danièle Blondel for kindly providing purified RV NCs. E.A.R. was supported by postdoctoral fellowships from the University Joseph Fourier and from the ANR program (ANR-06-JCJC-0126-01). A.A.V.A., F.C.A.G., and C.L. were supported by a fellowship from the french Ministère de l'Education Nationale, de la Recherche et de la Technologie. We thank the Partnership for Structural Biology for the excellent structural biology environment.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.09.042

References

1. Emerson, S. U. & Schubert, M. (1987). Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus. *Proc. Natl Acad. Sci. USA*, **84**, 5655–5659.
2. Kolakofsky, D., Le Mercier, P., Iseni, F. & Garcin, D. (2004). Viral RNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology*, **318**, 463–473.
3. Arnheiter, H., Davis, N. L., Wertz, G., Schubert, M. & Lazzarini, R. A. (1985). Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. *Cell*, **41**, 259–267.
4. Schoehn, G., Iseni, F., Mavrakis, M., Blondel, D. & Ruigrok, R. W. (2001). Structure of recombinant rabies virus nucleoprotein–RNA complex and identification of the phosphoprotein binding site. *J. Virol.* **75**, 490–498.
5. Iseni, F., Barge, A., Baudin, F., Blondel, D. & Ruigrok, R. W. (1998). Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures. *J. Gen. Virol.* **79**, 2909–2919.
6. Albertini, A. A., Clapier, C. R., Wernimont, A. K., Schoehn, G., Weissenhorn, W. & Ruigrok, R. W. (2006). Isolation and crystallization of a unique size category of recombinant rabies virus nucleoprotein–RNA rings. *J. Struct. Biol.* **158**, 129–133.
7. Albertini, A. A., Wernimont, A. K., Muziol, T., Ravelli, R. B., Clapier, C. R. & Schoehn, G. (2006). Crystal structure of the rabies virus nucleoprotein–RNA complex. *Science*, **313**, 360–363.
8. Green, T. J., Zhang, X., Wertz, G. W. & Luo, M. (2006). Structure of the vesicular stomatitis virus nucleoprotein–RNA complex. *Science*, **313**, 357–360.
9. Gérard, F. C. A., Ribeiro, E. A., Leyrat, C., Ivanov, I., Blondel, D. & Longhi, S. (2009). Modular organization of rabies virus phosphoprotein. *J. Mol. Biol.* **388**, 978–996.
10. Gérard, F., Ribeiro, E., Albertini, A., Zaccai, G., Ebel, C., Ruigrok, R. & Jamin, M. (2007). Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution. *Biochemistry*, **46**, 10328–10338.
11. Mavrakis, M., Mehouas, S., Real, E., Iseni, F., Blondel, D., Tordo, N. & Ruigrok, R. W. (2006). Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. *Virology*, **349**, 422–429.
12. Stillman, E. A. & Whitt, M. (1999). Transcript initiation and 5′-end modifications are separable events during vesicular stomatitis virus transcription. *J. Virol.* **73**, 7199–7209.
13. Chenik, M., Schnell, M., Conzelmann, K. K. & Blondel, D. (1998). Mapping the interacting domains between the rabies virus polymerase and phosphoprotein. *J. Virol.* **72**, 1925–1930.
14. Mavrakis, M., McCarthy, A. A., Roche, S., Blondel, D. & Ruigrok, R. W. (2004). Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. *J. Mol. Biol.* **343**, 819–831.
15. Kouznetzoff, A., Buckle, M. & Tordo, N. (1998). Identification of a region of the rabies virus N protein involved in direct binding to the viral RNA. *J. Gen. Virol.* **79**, 1005–1013.
16. Ribeiro, E. A., Jr, Favier, A., Gérard, F. C., Leyrat, C., Brutscher, B. & Blondel, D. (2008). Solution structure of the C-terminal nucleoprotein–RNA binding domain of the vesicular stomatitis virus phosphoprotein. *J. Mol. Biol.* **382**, 525–538.
17. Wyatt, P. J. (1998). Submicrometer particle sizing by multiangle light scattering following fractionation. *J. Colloid Interface Sci.* **197**, 9–20.
18. Tosatto, S. C., Bindewald, E., Hesser, J. & Manner, R. (2002). A divide and conquer approach to fast loop modeling. *Protein Eng.* **15**, 279–286.
19. Suhre, K. & Sanejouand, Y. H. (2004). ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.* **32**, W610–W614.
20. Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to

check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
21. Chen, R. & Weng, Z. (2002). Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins*, **47**, 281–294.
22. Chen, R., Li, L. & Weng, Z. (2003). ZDOCK: an initial-stage protein-docking algorithm. *Proteins*, **52**, 80–87.
23. Krol, M., Tournier, A. L. & Bates, P. A. (2007). Flexible relaxation of rigid-body docking solutions. *Proteins*, **68**, 159–169.
24. Jacob, Y., Real, E. & Tordo, N. (2001). Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains. *J. Virol.* **75**, 9613–9622.
25. Liang, S., Liu, S., Zhang, C. & Zhou, Y. (2007). A simple reference state makes a significant improvement in near-native selections from structurally refined docking decoys. *Proteins*, **69**, 244–253.
26. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, Palo Alto, CA.
27. Bradley, P., Misura, K. M. & Baker, D. (2005). Toward high-resolution de novo structure prediction for small proteins. *Science*, **309**, 1868–1871.
28. Toriumi, H. & Kawai, A. (2004). Association of rabies virus nominal phosphoprotein (P) with viral nucleocapsid (NC) is enhanced by phosphorylation of the viral nucleoprotein (N). *Microbiol. Immunol.* **48**, 399–409.
29. Toriumi, H., Honda, Y., Morimoto, K., Tochikura, T. S. & Kawai, A. (2002). Structural relationship between nucleocapsid-binding activity of the rabies virus phosphoprotein (P) and exposure of epitope 402-13 located at the C terminus. *J. Gen. Virol.* **83**, 3035–3043.
30. Ding, H., Green, T. J., Lu, S. & Luo, M. (2006). Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J. Virol.* **80**, 2808–2814.
31. Iseni, F., Baudin, F., Blondel, D. & Ruigrok, R. W. (2000). Structure of the RNA inside the vesicular stomatitis virus nucleocapsid. *RNA*, **6**, 270–281.
32. Albertini, A. A., Schoehn, G., Weissenhorn, W. & Ruigrok, R. W. (2008). Structural aspects of rabies virus replication. *Cell. Mol. Life Sci.* **65**, 282–294.
33. Green, T. J. & Luo, M. (2009). Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P. *Proc. Natl Acad. Sci. USA*, **106**, 11721–11726.
34. Thomas, D., Newcomb, W. W., Brown, J. C., Wall, J. S., Hainfeld, J. F., Trus, B. L. & Steven, A. C. (1985). Mass and molecular composition of vesicular stomatitis virus: a scanning transmission electron microscopy analysis. *J. Virol.* **54**, 598–607.
35. Blanchard, L., Tarbouriech, N., Blackledge, M., Timmins, P., Burmeister, W. P., Ruigrok, R. W. & Marion, D. (2004). Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. *Virology*, **319**, 201–211.
36. Jensen, M. R., Houben, K., Lescop, E., Blanchard, L., Ruigrok, R. W. & Blackledge, M. (2008). Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. *J. Am. Chem. Soc.* **130**, 8055–8061.
37. Curran, J. (1998). A role for the Sendai virus P protein trimer in RNA synthesis. *J. Virol.* **72**, 4274–4280.
38. Houben, K., Blanchard, L., Blackledge, M. & Marion, D. (2007). Intrinsic dynamics of the partly unstructured PX domain from the Sendai virus RNA polymerase co-factor P. *Biophys. J.* **93**, 2930–2940.
39. Bernard, C., Gely, S., Bourhis, J. M., Morelli, X., Longhi, S. & Darbon, H. (2009). Interaction between the C-terminal domains of N and P proteins of measles virus investigated by NMR. *FEBS Lett.* **583**, 1084–1089.
40. Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry*, **6**, 1948–1954.
41. Uversky, V. N. (1993). Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry*, **32**, 13288–13298.
42. Narayanan, T., Diat, O. & Bösecke, P. (2001). SAXS and USAXS on the high brilliance beamline at the ESRF. *Nucl. Instrum. Methods Phys. Res., Sect. A*, **467–468**, 1005–1009.
43. Konarev, P. V., Volkov, V. V., Sokolova, A., Koch, M. H. J. & Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282.
44. Semenyuk, A. V. & Svergun, D. (1991). GNOM—a program package for small-angle scattering data processing. *J. Appl. Crystallogr.* **24**, 537–540.
45. Zhang, C., Vasmatzis, G., Cornette, J. L. & DeLisi, C. (1997). Determination of atomic desolvation energies from the structures of crystallized proteins. *J. Mol. Biol.* **267**, 707–726.
46. Krippahl, L., Moura, J. J. & Palma, P. N. (2003). Modeling protein complexes with BiGGER. *Proteins*, **52**, 19–23.
47. Svergun, D., Barberato, C. & Koch, M. H. (1995). CRYSOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773.
48. Krol, M., Chaleil, R. A., Tournier, A. L. & Bates, P. A. (2007). Implicit flexibility in protein docking: cross-docking and local refinement. *Proteins*, **69**, 750–757.
49. Hess, B., Kutzner, C., Van der Spoel, D. & Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447.
50. Oostenbrink, C., Soares, T. A., van der Vegt, N. F. & van Gunsteren, W. F. (2005). Validation of the 53A6 GROMOS force field. *Eur. Biophys. J.* **34**, 273–284.
51. Oostenbrink, C., Villa, A., Mark, A. E. & van Gunsteren, W. F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**, 1656–1676.
52. Bussi, G., Donadio, D. & Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101.
53. Parrinello, M. & Rahman, A. (1981). Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182.
54. Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W. F. & Mark, A. E. (1999). Peptide folding: when simulation meets experiment. *Angew. Chem., Int. Ed.* **38**, 236–240.
55. Schuettelkopf, A. W. & van Aalten, D. M. F. (2004). PRODRG—a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 1355–1363.

SUPPLEMENTARY FIGURES

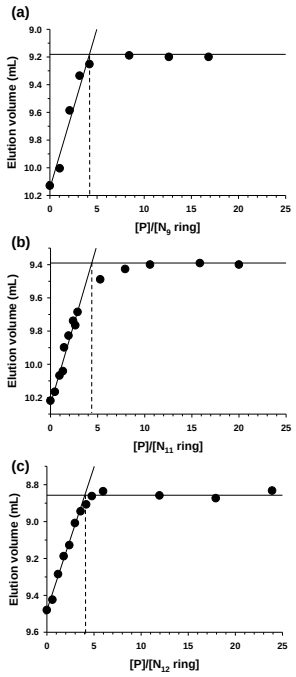


Figure S1. Titration of N_9-RNA (a), N_{11}-RNA (b) and N_{12}-RNA (c) with P. The concentration of N_m-RNA was kept constant between at 1 mM. The samples were separated on a S200 superdex column equilibrated in 20 mM Tris/HCl, 150 mM NaCl, pH 7.5. The maximum volume of the peak corresponding to the N_m-RNA-$(P_2)_2$ complexes from the titration series was plotted as a function of the [P]/[N_m-RNA] ratio.

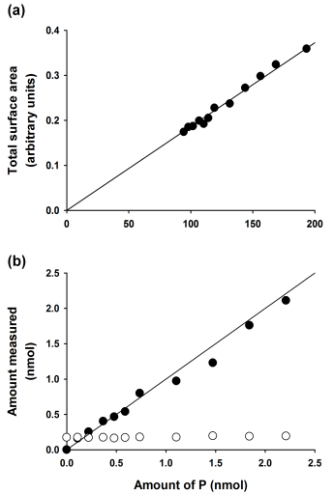


Figure S2. Deconvolution of the chromatographic series. (a) Total surface area versus total amount of added P and N_{10}-RNA. The line shows the linear regression (correlation coefficient = 0.992). **(b) Total amount measured versus total amount of added P.** The closed circles show the measured amount of P and the open circles show the sum of measured amounts of the N_{10}-RNA, N_{10}-RNA-(P_2) and N_{10}-RNA-$(P_2)_2$ complexes. The regression for P has a slope of 1.00 ± (correlation coefficient = 0.994).

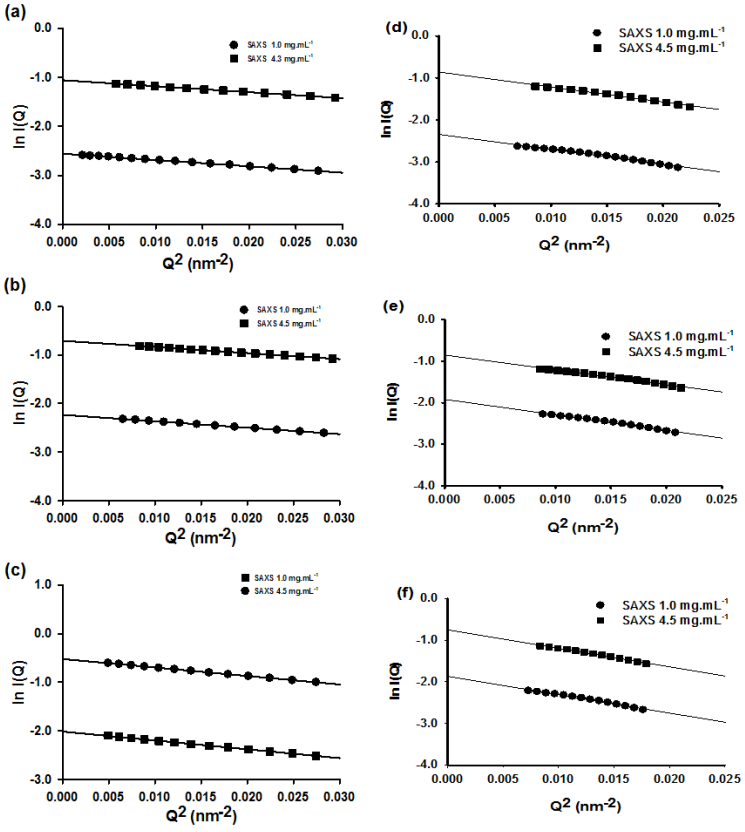


Figure S3. Guinier plots for the different N_m-RNA and N_m-RNA-$(P_{CTD})_2$ N_m-RNA-$(P_2)_2$ complexes. (a) N_{11}-RNA, (b) N_{11}-RNA-$(P_{CTD})_2$, (c) N_{11}-RNA-$(P_2)_2$, (d) N_{10}-RNA, (e) N_{10}-RNA-$(P_{CTD})_2$, (f) N_{10}-RNA-$(P_2)_2$.

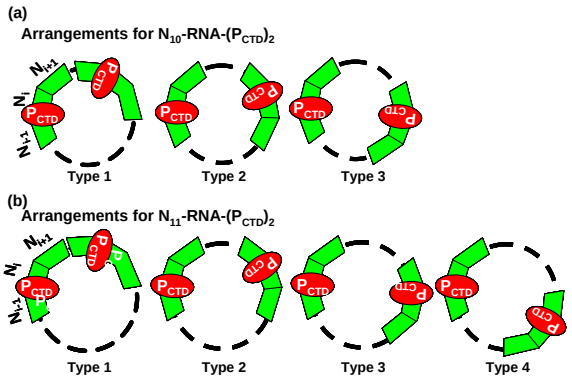
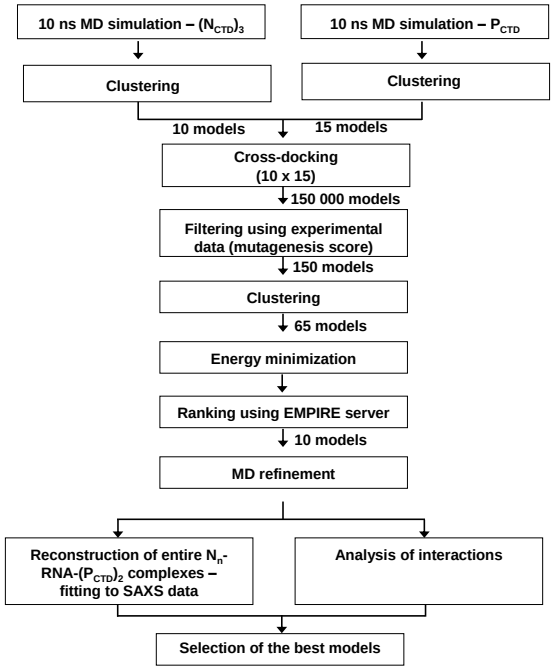


Figure S4. Reconstruction of the N_m-RNA complexes showing the different combinations used for replacing 3 N_{CTD} of the ring structure by the 3 N_{CTD} from the model generated in the presence of P_{CTD}. The black dots show the free N protomers, the green boxes show the 3 N protomers from the molecular docking. The P_{CTD} (in red) is located at the junction of the N_i and N_{i-1} protomers

(a)



(b)

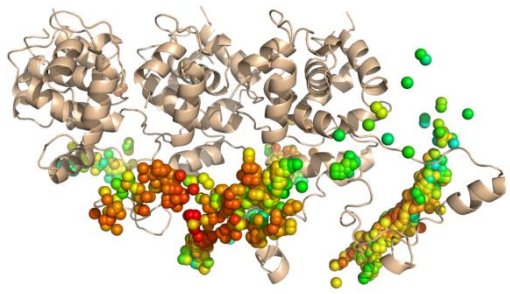


Figure S5. Flexible cross-docking procedure. (a) Scheme describing the flexible cross-docking procedure followed for generating models for the $(N_{CTD})_3$-P_{CTD} complex. (b) Mapping of the position of the centre of mass of P_{CTD} on the $(N_{CTD})_3$ complex in the 1000 decoys of one of the cross-docking step. Each bead represents the center of mass of one model and the beads are colour-coded from the lowest scores in blue to the highest scores in red.

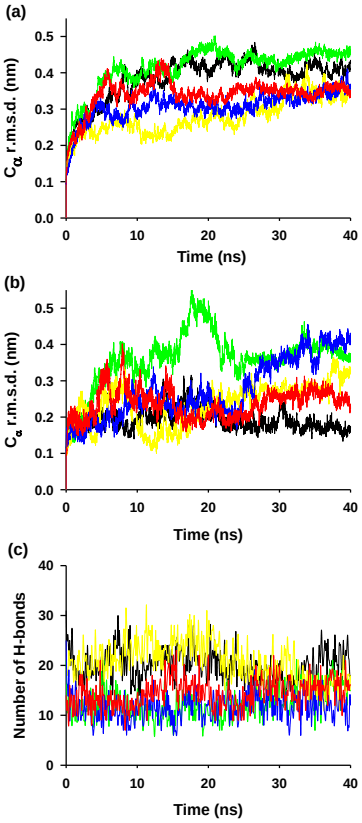


Figure S6. Molecular dynamics simulations. (a) Evolution of C_α r.m.s.d. for $(N_{CTD})_3$-P_{CTD} in the molecular dynamics simulations of the 5 best-scoring models. (b) **Evolution of C_α r.m.s.d. for P_{CTD} in the molecular dynamics simulations of the 5 best-scoring models. (c) Evolution of the number of intermolecular hydrogen bonds between $(N_{CTD})_3$-P_{CTD} and P_{CTD}.** After a quick equilibration (10 ns), the r.m.s.d. of the complex slowly drifts from 3.5 to 4.5 A (10ns < t < 40ns). During the simulations, the number of intermolecular hydrogen bonds fluctuated between roughly 10 and 28, suggesting that the structure is trapped in an energy minimum.

III. Données supplémentaires : Modélisation de l'interaction N-ARN-P

L'étude du complexe N_{10}-ARN-$(P_2)_2$ par SAXS et sa modélisation vise à obtenir un modèle structural atomique à la fois consistant avec la nature flexible et modulaire de P et les données de variation de contraste en SANS, qui fournissent une analyse à plus basse résolution de la structure du complexe (Euripedes de Almeida Ribeiro Jr, données non publiées). Le modèle doit également être en mesure d'expliquer les différences d'affinité observées par SPR entre la phosphoprotéine P_2, la délétion du domaine central $P_{\Delta CED}$, et le domaine P_{CTD} (Euripedes de Almeida Ribeiro Jr, données non publiées). Ces expériences indiquent une fixation plus forte de la phosphoprotéine entière comparée au domaine P_{CTD} seul, qui est susceptible de provenir de la fixation simultanée des deux P_{CTD} de chaque dimère à N_{10}-ARN, ou encore de changements complexes d'entropie conformationnelle entre les formes libres et liées de P.

A. Méthodes

Un ensemble de conformères de dimères de la phosphoprotéine du Virus de la Rage (P_2) dépourvus d'un de leurs domaines P_{CTD} a été construit en combinant les structures cristallographiques des domaines P_{CED} (pdb : 3L32) et P_{CTD} (pdb : 1VYI) avec des modélisations *ab initio* des régions désordonnées manquantes obtenues à l'aide du meta-serveur web LOMETS (Wu and Zhang 2007). Les conformations des deux segments N- et C-terminaux émergent de la région centrale de P_2 ont été échantillonnées à l'aide de méthodes de recherche conformationnelle aléatoire en utilisant les programmes AMMP (Harrison 1993) et VEGA ZZ (Pedretti, Villa et al. 2004). Alternativement, des structures correspondant à différents fragments de la phosphoprotéine ont été obtenues par dynamique moléculaire. Ces approches ont permis d'obtenir des modèles de P_2 ayant des rayons de giration (R_g) compris entre 3 et 8 nm. Cet ensemble a ensuite été utilisé en conjonction avec le modèle du complexe N_{10}-ARN-$(P_{CTD})_2$ obtenu précédemment (Ribeiro, Leyrat et al. 2009) afin de modéliser le complexe N_{10}-ARN-$(P_2)_2$. Pour ce faire, nous avons utilisé le logiciel SASREF (Petoukhov and Svergun 2005) permettant l'assemblage de corps rigides sur base de données SAXS tout en modifiant les paramètres du programme pour ne pas prendre en compte le profil SAXS expérimental (en augmentant artificiellement le poids des autres termes dans la fonction minimisée par l'algorithme génétique). Des conditions de contact entre les résidus connectant les domaines P_{CTD} présents sur le complexe N_{10}-ARN-$(P_{CTD})_2$ et les régions IDR_{CTDs} des phosphoprotéines délétées d'un de leurs deux domaines P_{CTD} ont été appliquées. Le programme SASREF a été lancé sur 5 cycles de l'algorithme uniquement (contre 100 normalement) de manière à obtenir un positionnement correct de la phosphoprotéine sur l'anneau N_{10}-ARN. Plusieurs ensembles de 15,000 structures ont été générés pour le complexe N_{10}-ARN-$(P_2)_2$ en répétant cette procédure avec différents ensembles de structures pour la protéine P_2. Ces modèles ont été utilisés comme pool de départ pour sélectionner un

sous-ensemble de structures de N_{10}-ARN-$(P_2)_2$ représentant les données expérimentales à l'aide du programme GAJOE (Genetic Algorithm for Judging the Optimization of Ensembles) qui fait partie du logiciel EOM (Bernado, Mylonas et al. 2007).

B. Modélisation du complexe N_{10}-ARN-$(P_2)_2$ en solution.

Etant donné que la phosphoprotéine est prédite comme possédant de longs segments désordonnés, la méthode d'optimisation d'ensembles (EOM) constitue l'approche de choix pour l'analyse des données SAXS du complexe N_{10}-ARN-$(P_2)_2$, car elle permet la représentation du complexe par un ensemble de structures dont la moyenne représente le profil expérimental et prend donc en compte la flexibilité de la phosphoprotéine. Les modèles produits pour le complexe N_{10}-ARN-$(P_2)_2$ sont constitués de deux dimères P_2 possédant un domaine de dimérisation et deux domaines P_{CTD} globulaires, connectés par deux segments désordonnés (IDR_{CTDs}), conformément aux prédictions bioinformatiques (Gérard, Ribeiro et al. 2009). Les régions N-terminales (résidus 1-89) sont également représentées par des chaînes désordonnées, alors que le modèle utilisé pour N_{10}-ARN-$(P_{CTD})_2$ est un élément fixe du système avec les deux P_{CTD} occupant des positions prédéfinies sur l'anneau N_{10}-ARN.

N-RNA complex	R_S (nm)	R_g (nm) (Guinier)	R_g (nm) (P(r))	<MM> (kDa)	$MM_{calc.}$ (kDa)	<Mw>/<Mn>
N_9-RNA	6.3 ± 0.1	6.2 ± 0.1		470 ± 10	483	1.005 ± 0.040
N_9-RNA-(P_2)	6.8 ± 0.1	N.D.		N.D.	550	N.D.
N_9-RNA-$(P_2)_2$	7.2 ± 0.1	N.D.		636 ± 10	617	1.001 ± 0.020
N_{10}-RNA	6.4 ± 0.1	6.1 ± 0.1		510 ± 7	537	1.001 ± 0.020
N_{10}-RNA-(P_2)	6.9 ± 0.1	N.D.		N.D.	604	N.D.
N_{10}-RNA-$(P_2)_2$	7.4 ± 0.1	7.2 ± 0.1		666 ± 9	671	1.000 ± 0.020
N_{11}-RNA	6.7 ± 0.1	6.3 ± 0.1		570 ± 10	590	1.002 ± 0.025
N_{11}-RNA-(P_2)	7.2 ± 0.1	N.D.		N.D.	657	N.D.
N_{11}-RNA-$(P_2)_2$	7.6 ± 0.1	7.4 ± 0.2		717 ± 8	724	1.000 ± 0.015
N_{12}-RNA	6.8 ± 0.1	N.D.	N.D.	650 ± 20	644	1.005 ± 0.050
N_{12}-RNA-(P_2)	7.2 ± 0.1	N.D.	N.D.	N.D.	711	N.D.
N_{12}-RNA-$(P_2)_2$	7.6 ± 0.1	N.D.	N.D.	N.D.	778	N.D.

Tableau 3 : Paramètres structuraux globaux obtenus par SEC-MALLS et SAXS pour les complexes circulaires N_m-RNA, N_m-RNA-(P_2) et N_m-RNA-$(P_2)_2$.

La corrélation obtenue entre l'ensemble sélectionné par EOM et le profil SAXS expérimental affiche une valeur de $\chi = 1.15$ (Figure 27a). La Figure 27b montre la distribution des valeurs de R_g dans le pool initial de structures et dans l'ensemble des conformères sélectionnés. Le R_g moyen pour l'ensemble sélectionné est de 7.2 nm et correspond bien à la valeur expérimentale obtenue par l'approximation de Guinier (Tableau 3). Les Figures 27c&e montrent l'arrangement des deux dimères de P attachés au complexe N_{10}-ARN. Les Figures 27d&f montrent une superposition des 20 modèles extraits de l'ensemble sélectionné

indiquant que P_2 occupent de multiples orientations par rapport au complexe N_{10}-ARN. Les valeurs de R_g des modèles présents dans le pool initial sont comprises entre 5.6 et 9.0 nm, alors que l'ensemble sélectionné contient des modèles dont les dimensions varient entre 6.5 et 8.1 nm, indiquant que les modèles générés à l'aide de notre procédure de modélisation couvre une gamme de R_g suffisamment large, et soulignant les importantes capacités d'extension et de contraction du complexe macromoléculaire formé entre la phosphoprotéine et l'anneau N_{10}-ARN. Les distances de 9 et 7 nm indiquées sur la Figure 27e correspondent à la distance moyenne observée dans l'ensemble sélectionné entre le centre de masse (COM) de l'anneau N_{10}-ARN et le COM des deux molécules P_2, et à la distance entre les COMs des deux molécules P_2, respectivement. Ces valeurs sont en accord avec les valeurs calculées à partir d'expériences de variation de contraste en diffusion de neutrons aux petits angles réalisées par Dr. Euripedes de Almeida Ribeiro Jr (Ribeiro et al., données non publiées).

Afin de tester la possibilité d'une fixation simultanée des quatre domaines P_{CTD} (2 par dimère) sur N-ARN, d'autres ensembles de structures ont été générés, notamment en imposant la fixation des quatre domaines P_{CTD} sur l'anneau N_{10}-ARN, et en utilisant de multiples simulations de dynamique moléculaire comportant des contraintes de distances entre les deux régions IDR_{CTDs} présentes sur chacune des chaînes du dimère, afin d'obtenir un bon échantillonnage conformationnel de la phosphoprotéine. Additionnellement des simulations de dimères P_2 tronqués en C-terminal au niveau du résidu 133 ont été réalisées, et les différents fragments ont été réassemblés dans SASREF *via* un protocole *rigid body* à cinq corps avec des contraintes de distance (complexe N_{10}-ARN-$(P_{CTD})_4$, $2x(IDR_{CTDs})_2$, et $2x(P\Delta C164)_2$). En effet, des mesures d'affinité réalisées par Dr. Euripedes de Almeida Ribeiro Jr suggèrent l'existence d'une seconde constante de dissociation entre les dimères de phosphoprotéine et les anneaux N_{10}-ARN, qui pourrait s'interpréter par la fixation simultanée des deux domaines P_{CTD} de chaque dimère de phosphoprotéine sur N_{10}-ARN, le second domaine P_{CTD} ayant une affinité plus faible (Ribeiro et al., données non publiées).

Cependant, et malgré la génération de plus de 30,000 modèles, je n'ai pas été en mesure d'obtenir une corrélation correcte avec les données expérimentales. Ces résultats indiquent une erreur importante dans les modèles générés, et ce en dépit d'un échantillonnage conformationnel large à la fois en termes de rayon de giration et de dimensions maximales des modèles (données non présentées), ce qui suggère que l'origine de la faible corrélation réside dans l'absence de certaines distances interatomiques caractéristiques au sein du complexe, qui pourraient correspondre à un ensemble de distances entre les P_{CTD} flexibles et l'anneau N_{10}-ARN qui sont absentes lorsque tous les P_{CTD} sont physiquement fixés sur N. De plus amples investigations seront nécessaires pour clarifier l'origine de cet apparent désaccord.

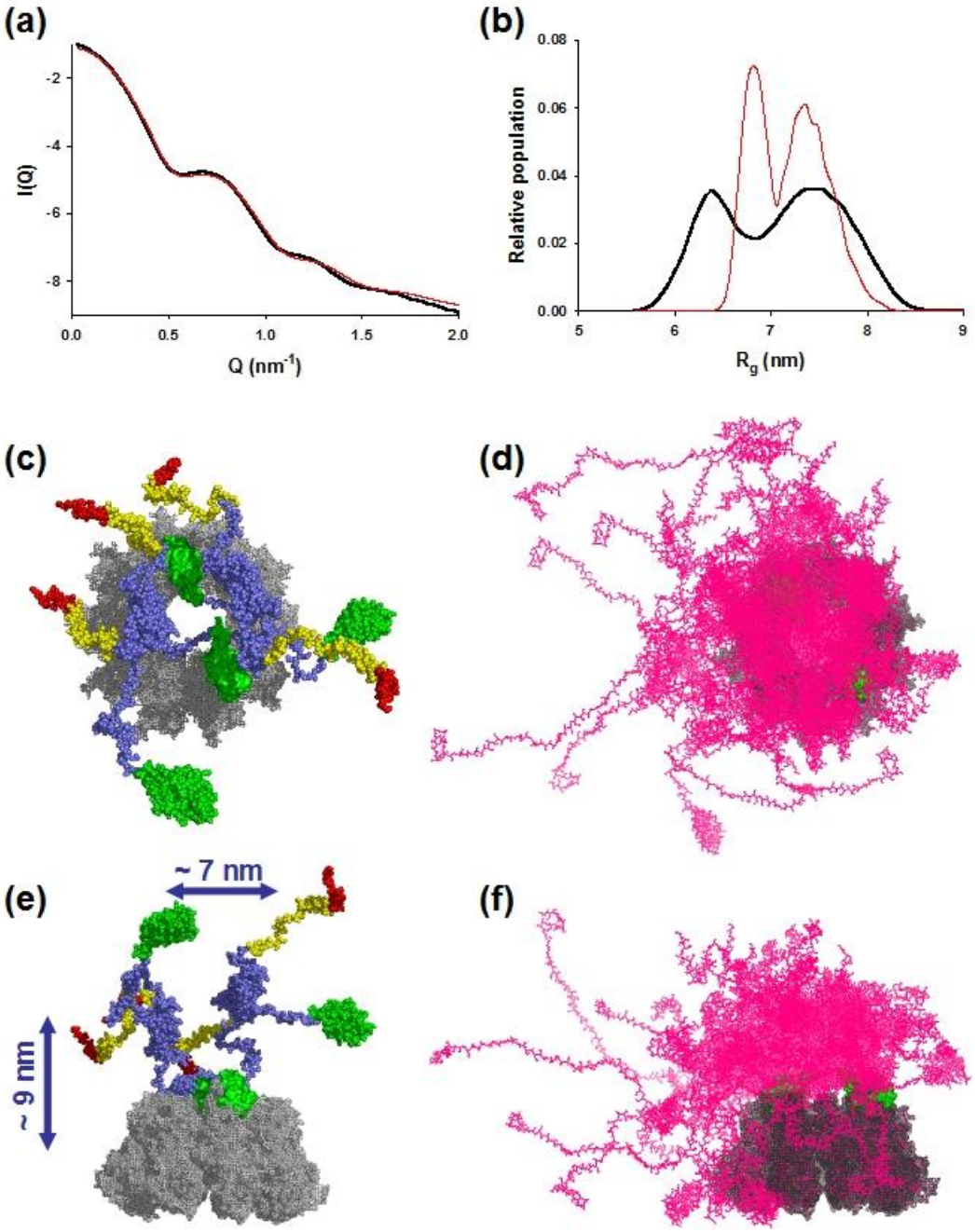


Figure 27 : Modélisation du complexe N_{10}-ARN-$(P_2)_2$. (a) Profiles de diffusion de rayons X aux petits angles (SAXS) pour le complexe N_{10}-ARN-$(P_2)_2$. La ligne noire représente les données expérimentales et la ligne rouge correspond à l'ensemble sélectionné par EOM. (b) Distribution des valeurs de R_g dans l'ensemble conformationnel sélectionné par EOM qui décrit le profile SAXS expérimental (ligne rouge) comparée à la distribution des valeurs de R_g dans le pool initial (ligne noire). (c, e) Conformère individuel de N_{10}-ARN-$(P_2)_2$ extrait de l'ensemble sélectionné par EOM. Les nucléoprotéines sont représentées en gris, les domaines PCTD en vert, les résidus 20 à 60 en jaune et les résidus 1 à 19 en rouge. (d, f) Superposition de 20 modèles extrait de l'ensemble sélectionné par EOM.

IV. Conclusion

Les données présentées dans ce chapitre nous permettent d'améliorer de manière significative notre compréhension et notre description de l'interaction entre les complexes nucléoprotéine-ARN, qui constituent la forme active du génome du Virus de la Rage et des autres virus à ARN négatifs, et la phosphoprotéine, qui joue notamment un rôle essentiel de cofacteur pour la polymérase virale L.

La structure modélisée du complexe N-ARN-P_{CTD}, obtenue grâce à une procédure de docking flexible, contrainte à l'aide de diverses données expérimentales, révèle l'interaction simultanée de P_{CTD} avec deux sous-unités du complexe N-ARN, et notamment l'implication des deux boucles C-terminales (C-loops) contenant l'unique site de phosphorylation de la nucléoprotéine localisé au niveau de la Ser389, et qui est capable de moduler l'affinité de P_{CTD} pour la nucléocapside (Toriumi and Kawai 2004). L'affinité de la forme phosphorylée du complexe nucléoprotéine-ARN pour la phosphoprotéine semble en effet être 1000 fois plus importante que pour la forme non phosphorylée (Ribeiro et al., données non publiées)

La structure cristallographique du complexe N_{10}-ARN-P_{CTD} chez le VSV (Green and Luo 2009) a été publiée peu de temps après la soumission de nos travaux (Ribeiro, Leyrat et al. 2009), et révèle un mode d'interaction similaire à ce qui a été prédit pour le Virus de la Rage (pour une comparaison, voir (Leyrat, Gerard et al. 2010) et Chap.VIII).

CHAPITRE VII: MODELISATION, PRODUCTION ET CARACTERISATION STRUCTURALE DU COMPLEXE N°P

gcq#63: "She Says She Can't Go Home Without a Chaperone"

(E. Costello)

I. Introduction

L'existence d'un complexe nucléoprotéine-phosphoprotéine N^0-P a été démontrée chez le VSV il y a plus de vingt ans (Davis, Arnheiter et al. 1986; Masters and Banerjee 1986; Peluso and Moyer 1988). Ce complexe constitue la forme soluble de la nucléoprotéine (N^o) dans les cellules infectées. La phosphoprotéine fonctionne comme une chaperonne pour la nucléoprotéine et évite sa fixation non-spécifique aux ARN cellulaires ou aux ARN messagers viraux. Cependant, ces travaux pionniers n'ont pu définir clairement la stœchiométrie du complexe, ni identifier les régions des protéines N et P impliquées dans l'interaction. Des études plus récentes réalisées sur le complexe N^0-P du RV (Mavrakis, Iseni et al. 2003; Mavrakis, Mehouas et al. 2006), mais également du VSV (Chen, Ogino et al. 2007), suggèrent une stœchiométrie 1 : 2 (ratio N/P) pour le complexe, et ont démontré que la région N-terminale de la phosphoprotéine (résidus 4 à 60 chez RV, les résidus 14 à 40 étant les plus critiques ; résidus 1 à 40 ou 11 à 30 chez VSV) constituait l'élément essentiel responsable de la fonction de chaperonne de N.

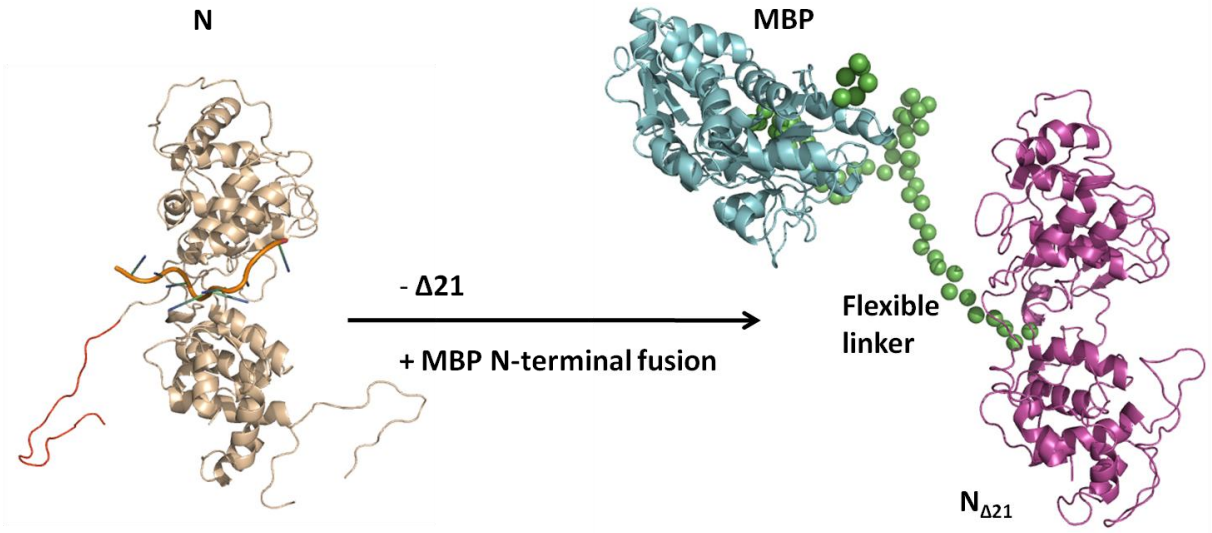


Figure 28 : **Illustration des constructions utilisées pour la production des complexes VSV $N_{Δ21}°P_{NTD}$ et RV $N_{Δ23}{}^0P_{NTD}$.** Les protéines ont été clonées dans un vecteur pET-M40 (EMBL vectors) et exprimées dans *E.coli*.

Les tentatives pour coexprimer N et P et purifier le complexe N^0-P ou un complexe N^0-P_{NTD} après protéolyse ménagée ont montré des rendements faibles qui ne permettaient pas d'envisager d'études structurales.

Dans un premier temps, la structure du complexe a été modélisée en utilisant une procédure d'arrimage moléculaire basée sur l'utilisation de modèles atomiques prédits par modélisation *ab initio* de la région N-terminale de la phosphoprotéine et des structures

cristallographiques disponibles pour les protéines N de RV et VSV (Albertini, Wernimont et al. 2006; Green, Zhang et al. 2006). Ce travail est résumé dans la deuxième partie du chapitre.

Dans un second temps, une nouvelle stratégie expérimentale pour la production du complexe N^0-P a été développée. Pour remédier au problème de production de N^0-P, nous avons envisagé de produire la protéine N sous la forme d'un complexe N-ARN qui peut être produit en grandes quantité et ensuite de déplacer l'ARN avec la protéine P entière ou avec P_{NTD}. Toutefois, avec la protéine N sauvage, l'addition de PNTD ou de P entière n'a pas permis de déplacer l'ARN. Dans le complexe N-ARN, l'assemblage des molécules de N est stabilisé par le bras N-terminal (aa 1-21 chez VSV, aa 1-23 chez RV) et par la boucle C-terminale (aa 340-354 chez VSV, aa 353-363 chez RV) qui s'échangent entre protomères voisins (Albertini, Wernimont et al. 2006; Green, Zhang et al. 2006; Luo, Green et al. 2007). Nous avons donc imaginé de déstabiliser le complexe en délétant le bras N-terminal. La délétion des 21 ou 23 premiers résidus de la nucléoprotéine respectivement du VSV ou du RV (Figure 28) a rendu l'encapsidation de l'ARN réversible en présence de P_{NTD} et a permis de produire et de purifier des quantités importantes de complexes VSV $N_{\Delta 21}{}^0P_{NTD}$ et RV $N_{\Delta 23}{}^0P_{NTD}$ en vue de leur caractérisation structurale. Le travail a été réalisé avec la précieuse collaboration de Filip Yabukarski, stagiaire de Master2 avec lequel j'ai travaillé sur ce projet. Ces travaux sont présentés dans les troisième et quatrième parties de ce chapitre sous la forme d'un article accompagné de données supplémentaires.

II. Prélude: Modélisation ab initio du complexe N^0-P

Les premiers travaux réalisés sur le complexe N^0-P à mon arrivée au laboratoire ont consistés à modéliser le complexe à l'aide des méthodes de bioinformatique structurale utilisant la dynamique moléculaire (Adcock and McCammon 2006), l'arrimage moléculaire (Chen, Li et al. 2003; Gray, Moughon et al. 2003; Krippahl, Moura et al. 2003; Wang, Bradley et al. 2007), les méthodes de classement empiriques des modèles basées sur la conservation (Duan, Reddy et al. 2005), les caractéristiques physiques statistiquement observées dans les interfaces observées dans la PDB (Bernauer, Aze et al. 2007), ou encore les potentiels d'énergie pseudo-empiriques (Camacho and Zhang 2005; Champ and Camacho 2007; Liang, Liu et al. 2007; Liang, Wang et al. 2008), ainsi que l'analyse des résidus constituant les « hot spots » de l'interface (Kortemme, Kim et al. 2004; Schymkowitz, Borg et al. 2005; Darnell, LeGault et al. 2008)). Ces calculs m'ont permis de mieux appréhender le système N^0-P avant d'effectuer tout travail expérimental.

A. Procédure de modélisation du complexe N^0-P_{NTD}.

Le complexe N^0-P a été modélisé en appliquant la procédure présentée dans la Figure 29. La structure des protomères de nucléoprotéine de RV et VSV ont été extraites de leurs structures cristallographiques respectives sous la forme de complexes N-ARN (Albertini, Wernimont et al. 2006; Green, Zhang et al. 2006). La boucle manquante du domaine C-terminal de RV N a été modélisée comme décrit dans (Ribeiro, Leyrat et al. 2009). Un modèle pour la nucléoprotéine de Chandipura virus a été construit par homologie *via* le serveur SWISS-MODEL (Arnold, Bordoli et al. 2006). En ce qui concerne P_{NTD} (VSV P_{NTD} correspondant aux aa 1-60 de la souche Orsay du sérotype Indiana, RV P_{NTD} correspondant aux aa 4-60 de la souche CVS, CHANV P_{NTD} correspondant aux aa 1-60), un ensemble de structures a été généré en soumettant la séquence d'acides aminés au méta-serveur LOMETS (Wu and Zhang 2007). Cette procédure nous a permis d'obtenir 5 modèles non-redondants pour VSV P_{NTD}, 9 pour RV P_{NTD} et 4 pour CHANV P_{NTD}. La qualité des modèles obtenus a été vérifiée avec PROCHECK (Laskowski, Macarthur et al. 1993).

Dans un premier temps, une procédure d'arrimage global à l'aide du logiciel BiGGER a été utilisée pour générer des ensembles de modèles du complexe N^0-P_{NTD} (Palma, Krippahl et al. 2000; Krippahl, Moura et al. 2003). Ces modèles ont ensuite été regroupés en sous-ensembles (« *clusters* ») en utilisant un seuil à 2 Å et les 100 meilleurs modèles de chaque calcul, classés en fonction du score électrostatique de BiGGER, ont été soumis au protocole d'arrimage local et d'affinement de RosettaDock (Wang, Schueler-Furman et al. 2007). Les modèles résultants ont été minimisés en énergie dans GROMACS (van der Spoel, Lindahl et al. 2005) en utilisant la méthode de la plus grande pente (*steepest descent*) et la méthode des

gradients conjugués (*conjugate gradients*) avec le champ de force OPLS-AA (Kaminski, Friesner et al. 2001). Les modèles ont ensuite été classés à l'aide d'un filtre de conservation calculé pour chaque résidu d'après des alignements de séquences des différents *Vesiculovirus* et *Lyssavirus* avec le programme AL2CO (Pei and Grishin 2001). Les résidus présents à l'interface ont été identifiés en utilisant le logiciel PSAIA (Mihel, Sikic et al. 2008), et un score de conservation moyen a été calculé pour les résidus de cette interface (Duan, Reddy et al. 2005). Les 25% de modèles affichant les scores de conservation moyens les plus bas ont été éliminés, et les modèles restants ont été reclassés en utilisant la fonction AffinitySCORE (Audie and Scarlata 2007; Audie 2009).

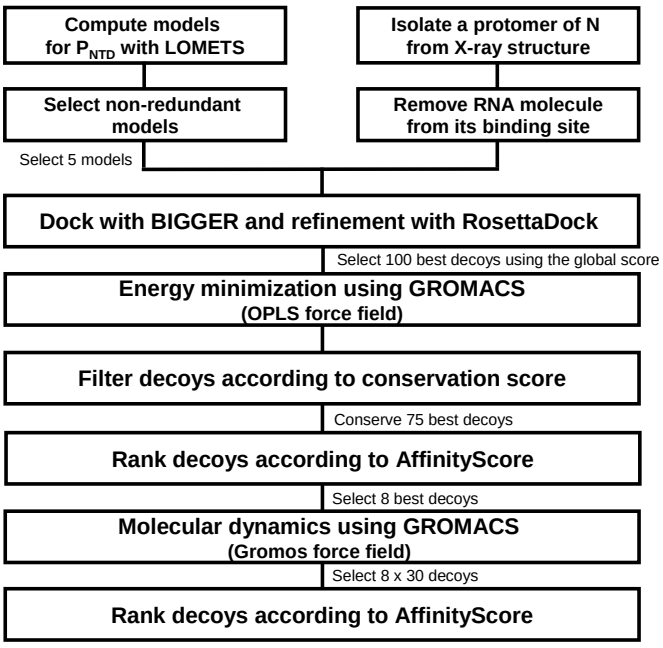


Figure 29 : Stratégie utilisée pour la modélisation structurale du complexe N^0-P_{NTD}. Schéma décrivant les différentes étapes utilisées dans la procédure de cross-docking flexible pour générer des modèles du complexe N^0-P_{NTD}.

Nous avons ensuite utilisé une procédure d'affinement flexible des modèles par dynamique moléculaire, cette approche ayant été décrite comme améliorant substantiellement la qualité des modèles obtenus et leur discrimination à l'aide de fonctions d'énergie d'interaction empiriques (Krol, Chaleil et al. 2007; Krol, Tournier et al. 2007). Pour chaque système viral, les 8 modèles les mieux classés ont été simulés dans GROMACS 4.0.4 en solvant explicite avec le champ de force GROMOS 53a6 (Oostenbrink, Villa et al. 2004; Oostenbrink, Soares et al. 2005). Les modèles du complexe ont été immergés dans une boîte de molécules d'eau SPC/E (Berendsen, Grigera et al. 1987), avec une distance minimum entre le soluté et la boîte de 0.9 nm. La charge négative nette du système a été neutralisée par addition du nombre d'ions Na^+ requis. Les contributions électrostatiques longue distance ont été calculées par sommation d'Ewald (particle-mesh Ewald, PME). Les atomes d'hydrogène ont été traités en tant que sites virtuels, permettant ainsi l'utilisation d'un pas d'intégration de 5 femtosecondes. Le thermostat « *v-rescale* » (Bussi, Donadio et al. 2007) et le barostat Parrinello-Rahman (Parrinello and Rahman 1981) ont été utilisés pour maintenir une

température de 300 K et une pression de 1 atm. Après 2500 étapes de minimisation d'énergie, le système a été équilibré pendant 125 ps avec les atomes de protéine restreints. Des simulations de 30 ns ont ensuite été calculées et les trajectoires résultantes analysées avec GROMACS afin d'obtenir l'évolution temporelle de plusieurs paramètres tels que le nombre de liaisons hydrogènes interchaînes, ainsi que le RMSD des Cα des complexes entiers et des ligands P_{NTD} (Figure 30). Enfin, des structures extraites à intervalles régulier de 1 ns au cours des trajectoires ont été classés à l'aide du logiciel AffinitySCORE (Audie and Scarlata 2007; Audie 2009). Les meilleurs modèles ont été analysés avec PISA et PIC (Krissinel and Henrick 2007; Tina, Bhadra et al. 2007) et les figures réalisées dans PyMOL (DeLano 2002).

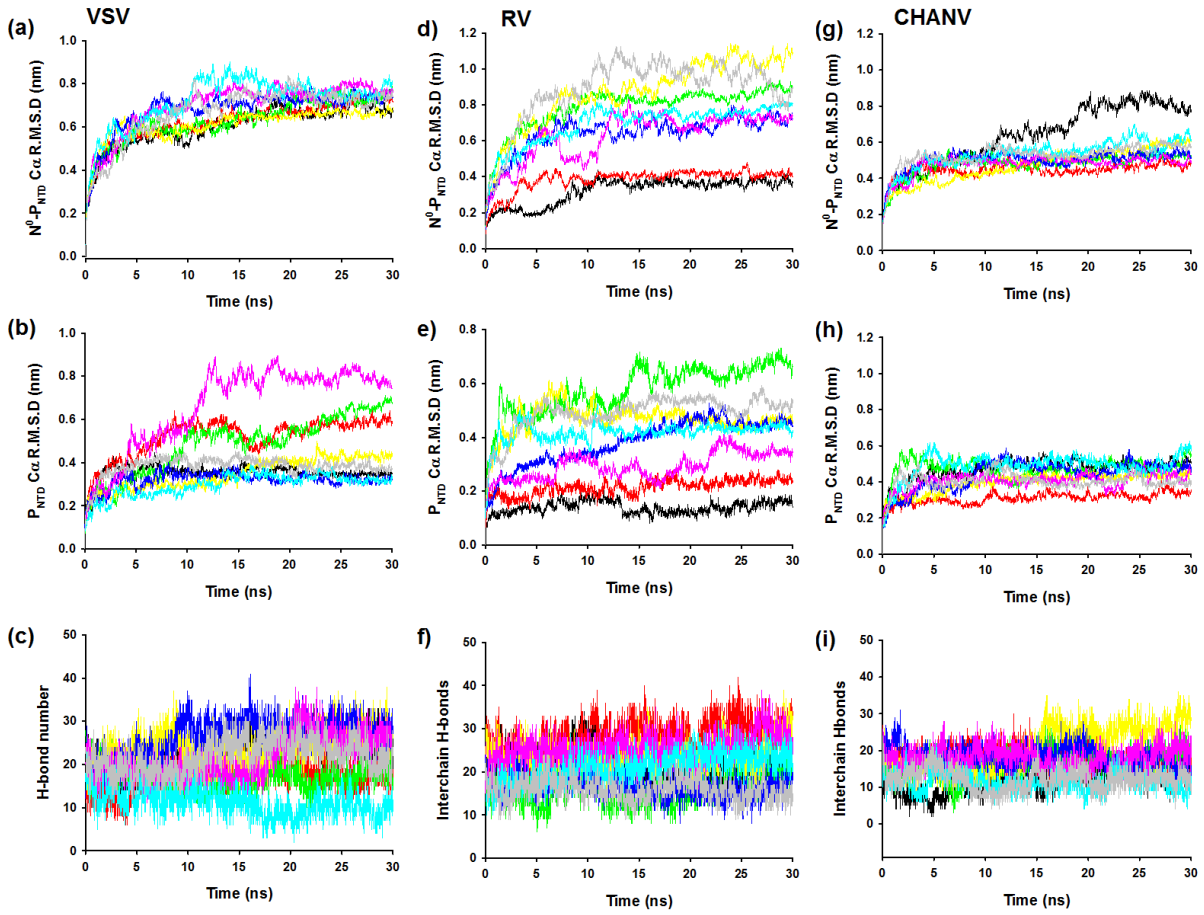


Figure 30 : Simulations de Dynamique Moléculaire avec VSV (a, b, c), RV (d, e, f) CHANV (g, h, i). (a, d, e) Evolution du Cα r.m.s.d. pour N^0-P_{NTD} dans les simulations de dynamique moléculaire des 8 meilleurs modèles. Après une rapide équilibration (~10 ns), le r.m.s.d. du complexe tend à atteindre un plateau caractéristique d'une structure équilibrée pour la plupart des modèles simulés. **(b, e, h) Evolution du Cα r.m.s.d. pour P_{NTD} dans les simulations de dynamique moléculaire des 8 meilleurs modèles.** Ici encore, le r.m.s.d arrive à un plateau après un temps d'équilibration d'environ 10ns. Pour certains modèles, la valeur se stabilise à un r.m.s.d. pouvant atteindre 0.8 nm, reflétant d'importants changements conformationnels pour P_{NTD}. **(c, f, i) Evolution du nombre de liaisons hydrogènes intermoléculaires entre N^0 et P_{NTD}.** Durant les simulations, le nombre de liaisons hydrogènes intermoléculaires (cut-off radial = 0.35nm, cut-off angulaire = 30°) fluctue entre environ 5 et 40 sur l'ensemble des modèles étudiés, et tend généralement à se stabiliser dans la partie équilibrée de la trajectoire.

B. Structure prédite du complexe N^0-P_{NTD}.

La modélisation du complexe N^0-P a été réalisée sur 3 systèmes homologues, VSV, Chandipura virus (CHANV) et le virus de la rage (RV) et avec de multiple conformations de P_{NTD} obtenues *via* le serveur de metathreading LOMETS (Wu and Zhang 2007). Dans un premier temps, une recherche globale par arrimage rigide avec flexibilité implicite des chaînes latérales à l'aide du programme BiGGER a été réalisée en utilisant chacun des différents modèles non-redondants de P_{NTD} et la structure de N^0. Ce calcul a révélé que la plupart des modèles possédant des scores élevés avait une molécule de P_{NTD} localisée dans la région de la cavité de fixation à l'ARN de N. La Figure 31a montre le centre de masse de P_{NTD} sur la structure du protomère de N^0 dans les 100 modèles les mieux classés obtenus pour VSV avec chacun des 5 modèles initiaux de P_{NTD}. Un classement similaire en sous-ensembles des meilleures solutions montrant P_{NTD} inséré dans la cavité de fixation à l'ARN de N^0 a été obtenu avec les modèles générés pour RV et CHANV (Figures 32a&c). Après minimisation d'énergie, les modèles ont été classés en utilisant le logiciel AffinityScore (Audie and Scarlata 2007; Audie 2009). Dans les trois systèmes viraux étudiés, les premiers 40 aa de P_{NTD} partagent un plus grand nombre de contacts avec N^0 que le reste de la molécule, en accord avec les données expérimentales indiquant que les résidus 11-30 de VSV P_{NTD} sont les plus critiques pour l'interaction avec N^0 (Chen, Ogino et al. 2007) et que les résidus 4-40 de RV P_{NTD} sont résistants à la digestion dans des expériences de protéolyse limitée du complexe N^0-P (Mavrakis, Mehouas et al. 2006). Les 8 meilleurs modèles ont été soumis à 30 ns de simulations de dynamique moléculaire, et des images extraites à chaque nanoseconde à partir des 8 trajectoires ont été classées en utilisant la fonction AffinityScore. Dans le modèle obtenant le meilleur score pour VSV, P_{NTD} est inséré dans la cavité de fixation à l'ARN de N (Figure 31b) et la remplit totalement (Figure 31c). P_{NTD} adopte une conformation triangulaire (Figure 31d). Les sous-domaines N- et C-terminaux de N se replient sur le cœur de N pendant la simulation, et le premier côté de P_{NTD} (aa 1-17) forme une hélice α (aa 3-14) qui contacte le sous-domaine C-terminal de N (Figure 31b). Le second côté (aa 18-38) s'insère profondément dans la cavité de fixation à l'ARN. La région comprenant les résidus 23 à 29 forme une hélice α qui persiste le long de la simulation, mais s'étend transitoirement jusqu'au résidu 37. Le troisième côté du triangle (aa 39-60) se replie à l'avant de la cavité de fixation à l'ARN de N, et son extrémité C-terminale (aa 52-57) forme une courte hélice α qui interagit avec le premier côté en formant un petit cœur hydrophobe (L4, V7, L11, Y53, F54, A57) (Figure 31d).

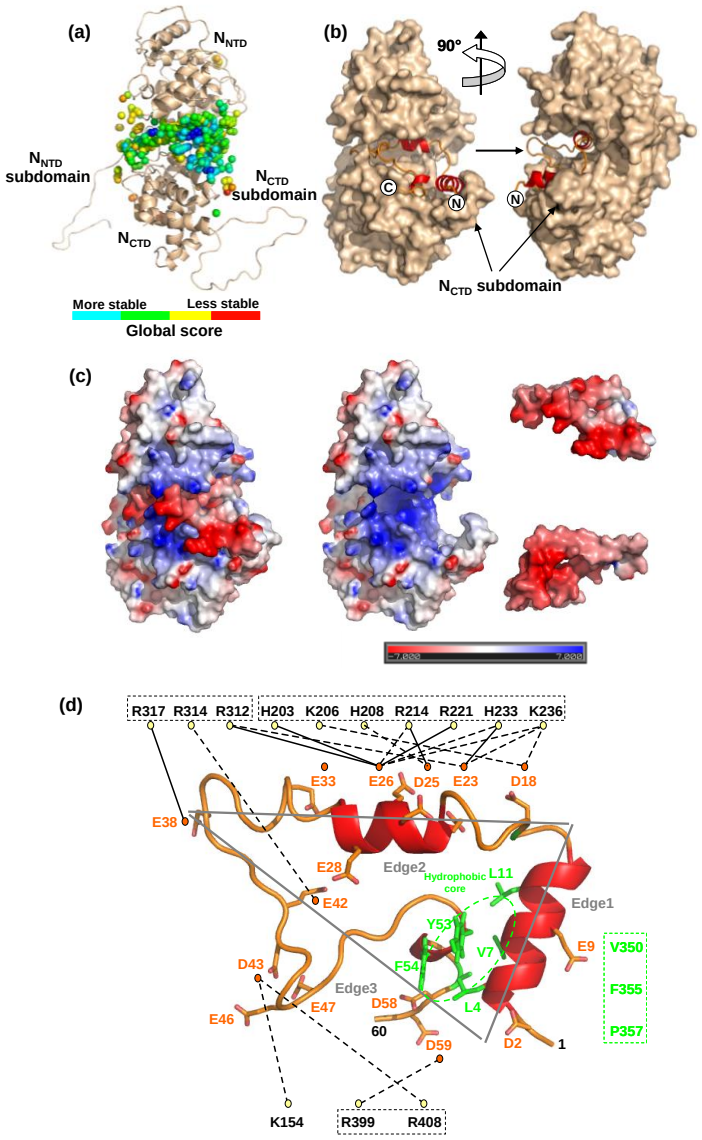


Figure 31 : Modélisation moléculaire du complexe N^0-P_{NTD}. (a) Localisation du centre de masse de P_{NTD} sur la structure de N^0 pour les modèles les mieux classés obtenus dans la première étape de la procédure de docking. **(b) Structure prédite du complexe formé entre N^0 et P_{NTD}.** P_{NTD} est représenté en cartoons sur la surface de N^0. Les segments hélicoïdaux de P_{NTD} sont représentés en rouge. **(c) Représentation en surface du complexe N^0-P_{NTD}** montrant l'occupation de la cavité de liaison à l'ARN de N^0 par P_{NTD}, et colorée d'après le potentiel électrostatique de N^0 et P_{NTD}. La structure de P_{NTD} montrée sur la droite a été extraite de la structure du complexe N^0-P_{NTD} et est dessinée dans deux orientations, celle du haut montrant la face exposée au solvant et celle du bas la face enfouie dans N^0. **(d) Représentation schématique de P_{NTD} dans le complexe N^0-P_{NTD},** soulignant les résidus chargés négativement et les résidus hydrophobes formant un cœur intramoléculaire. Les résidus de N^0 impliqués dans des ponts salins avec des résidus de P_{NTD} sont indiqués et encadrés au-dessus et en-dessous de la structure. Les lignes en gras indiquent des ponts salins. Les lignes en pointillés montrent les contacts à < 0.6 nm entre résidus chargés dans N^0 et P_{NTD}. Les résidus en vert encadrés sur la droite sont les résidus hydrophobes de N^0 formant un cœur intermoléculaire hydrophobe.

Un paramètre de base permettant de caractériser l'interaction entre deux protéines est la taille de l'interface, calculée comme la somme de la surface accessible au solvant (ASA) des partenaires isolés moins l'ASA du complexe. Pour le complexe N^0-P_{NTD}, l'aire de l'interface est 5,780 $Å^2$, correspondant à l'enfouissement de 3,080 $Å^2$ de P_{NTD} (53 % de son ASA totale dans sa conformation liée) et de 2,700 $Å^2$ de N^0 (12 % de son ASA totale). La formation du complexe implique des interactions entre 45 résidus de N et 37 résidus de P_{NTD}. La taille importante de l'interface et le grand nombre de résidus en interaction suggérait une affinité élevée (ΔG_{bind} = - 35.8 kcal/mol prédite par la fonction AffinityScore pour le modèle le mieux classé). Les interactions hydrophobes sont limitées à un petit cœur hydrophobe formé par les résidus V7, F10 et F14 dans l'hélice N-terminal de P_{NTD} et les résidus V350, F355 et P357 dans le sous-domaine C-terminal de N^0. Cependant, le complexe est stabilisé par un grand nombre d'interactions polaires. Une analyse fournie par le serveur PISA indique la présence de 44 liaisons hydrogènes intermoléculaires. La cavité de fixation à l'ARN possède un potentiel électrostatique hautement positif (14 résidus positivement chargés), alors

que P_{NTD} est fortement chargé négativement (15 résidus négativement chargés) (Figure 31c). La structure du modèle sélectionné contient 6 ponts salins impliquant 6 résidus positivement chargés de N et 4 résidus négativement chargés de P_{NTD} (Figure 31d). De plus, il existe 12 autres contacts entre résidus de charges opposées localisés à moins de 0.6 nm l'un de l'autre mais qui ne sont pas considérés comme ponts salins par le serveur PISA (Figure 31d).

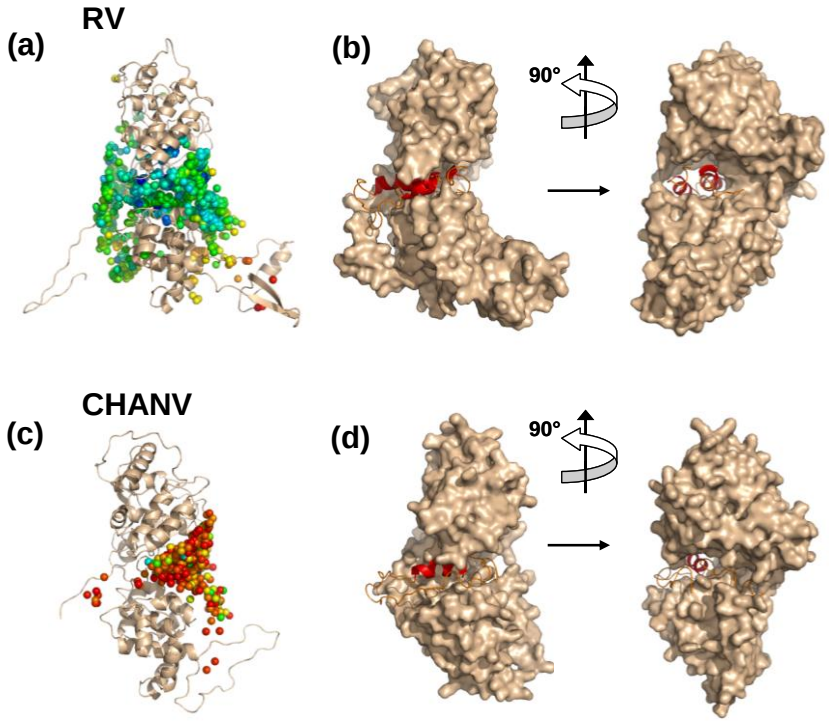


Figure 32 : Structure du complexe N^0-P_{NTD} généré par la procédure de docking pour RV et CHANV. (a) et (c) Localisation du centre de masse de P_{NTD} sur la structure de N^0 pour les 100 modèles les mieux classés pour chaque modèle initial de P_{NTD} obtenus dans la première étape de la procédure de docking. Chaque sphère représente le centre de masse d'un modèle et est colorée en fonction du score électrostatique de BIGGER à partir des scores les plus bas en rouge aux scores les plus élevés en bleu. **(b) and (d) Structure du modèle le mieux classé obtenu à la fin de la procédure de docking.**

Comme la région P_{NTD} se lie à N^0 (Mavrakis, Mehouas et al. 2006; Chen, Ogino et al. 2007), la présence de structures hélicoïdales résiduelles et l'existence d'un ensemble de conformères compactes (Chap.III) suggéraient que P_{NTD} puisse se replier en se liant à son partenaire. Le modèle généré pour le complexe N^0-P_{NTD} par docking moléculaire révèle la possible formation d'un réseau extensif d'interactions polaires et hydrophobes entre des résidus de la cavité de fixation de l'ARN de N^0 et de nombreux résidus dans P_{NTD}. Cependant, le repliement de P_{NTD} dans ce complexe est peu stabilisé par des interactions intramoléculaires longues distances, à l'exception de la formation d'un petit core hydrophobe entre les parties N- et C-terminales de P_{NTD}. Les interactions intermoléculaires avec N^0 pourraient fournir suffisamment d'énergie pour stabiliser P_{NTD} dans une conformation bien définie. Le manque d'interactions intramoléculaires au sein de P_{NTD} explique l'absence de structure repliée à l'état isolé, et fournit un mécanisme simple permettant d'assurer que la région P_{NTD} soit désordonnée en l'absence de son partenaire. L'aire de l'interface calculée pour le complexe N^0-P_{NTD} est de 5780 $Å^2$, ce qui constitue une valeur élevée comparée à une interface protéine-protéine typique (Jones and Thornton 1996), et même comparée à des interfaces observées entre protéines ordonnées et désordonnées (Mészáros, Tompa et al. 2007). Dans une étude portant sur 75 complexes protéine-protéine, la surface totale masquée au solvant par l'interaction entre les deux partenaires variait entre 1600 et 4660 $Å^2$ (Lo Conte, Chothia et al. 1999). Le plus souvent, les complexes ayant des interfaces de taille supérieure à 2000 $Å^2$

impliquent une transition ordre-désordre ou des changements de conformation importants pour au moins l'une des deux protéines (Lo Conte, Chothia et al. 1999), comme cela est observé ici pour P_{NTD}. Le complexe N^0-P_{NTD} est stabilisé principalement par des interactions polaires, mais encore une fois le nombre de liaisons hydrogènes et de ponts salins excède largement le nombre typiquement retrouvé dans d'autres complexes protéine-protéine (Mészáros, Tompa et al. 2007). Ces caractéristiques résultent en une énergie d'interaction prédite de -35.8 kcal/mol avec la fonction AffinityScore (Audie 2009). Des valeurs similairement élevées ont été obtenues avec d'autres fonctions (Liang, Liu et al. 2007). Ces valeurs anormalement élevées d'énergie libre d'interaction sont surement largement surestimées car le calcul suppose une interaction de type corps rigides et estime la somme des interactions stabilisantes et déstabilisantes sans prendre en compte le changement d'entropie conformationnelle associé à la transition désordre-ordre de l'un des interactants. Le mécanisme de repliement induit par l'interaction avec un partenaire a été proposé comme un mécanisme fournissant une grande spécificité sans le corollaire d'une haute affinité (Wright and Dyson 2009). Au cours du cycle de réplication viral, le complexe N^0-P ne se forme que transitoirement, en tant qu'intermédiaire dans la synthèse des nouvelles nucléocapsides. La chaperonne P doit être remplacée par l'ARN génomique, et l'affinité de P pour N^0 ne peut donc pas être trop élevée.

La caractérisation expérimentale du complexe N^0-P_{NTD} et notamment la résolution de sa structure cristallographique chez VSV fournissent un modèle de l'interaction entre N^0 et P_{NTD} qui est sensiblement différent de celui prédit en utilisant l'approche *ab initio*. En effet, le modèle théorique possède une interface étendue et enrichie en interactions polaires, alors que la structure expérimentale fait au contraire intervenir un grand nombre d'interactions hydrophobes (cf. partie III, ARTICLE VII). Cette large interface couplée au grand nombre d'interactions polaires résulte, de manière assez similaire à l'erreur observée pour la modélisation *ab initio* du domaine P_{CED} de RV (cf. INTERLUDE), d'un biais introduit dans les premières étapes du calcul. En effet, le choix du score électrostatique de BiGGER pour le classement primaire des modèles a artificiellement favorisé les modèles ayant une grande surface d'interaction avec la cavité centrale de N, car ils présentent une neutralisation maximale des charges de P_{NTD} et de N (Figure 31c). Cette erreur s'est par la suite répercutée sur toute la procédure, empêchant ainsi la prédiction d'une petite surface de contact hydrophobe localisée sur la face 3' de la cavité de fixation à l'ARN de N.

Malgré ces imprécisions, il est intéressant de noter que les récents progrès dans les méthodes de modélisation *ab initio* (Zhang 2008) et de docking protéine-protéine (Vajda and Kozakov 2009; Zacharias 2010) nous ont permis de produire une prédiction grossièrement correcte du site d'interaction, et ce malgré l'absence d'information concernant la conformation de P_{NTD} dans le complexe. Par ailleurs, les meilleurs modèles obtenus pour le VSV et CHANV suggèrent une interaction entre P_{NTD} et la région de la boucle C-terminale impliquée dans la formation de contacts N-N dans les complexes N-ARN. Cette interaction, qui est également suggérée par d'autres résultats obtenus chez VSV (cf. partie III, ARTICLE VII), n'est pas présente dans les modèles issus de l'arrimage rigide, mais apparaît spontanément au cours de certaines des simulations de dynamique moléculaire.

III. ARTICLE VII: Structural Basis for the Chaperone Activity of the Vesicular Stomatitis Virus Phosphoprotein

Auteurs: Cédric Leyrat*, Filip Yabukarski*, Nicolas Tarbouriech, Euripedes de Almeida Ribeiro Jr, Francine Gérard, Rob W. H. Ruigrok, and Marc Jamin.

* Cédric Leyrat et Filip Yabukarski ont contribués de manière équivalente

En préparation

STRUCTURAL BASIS FOR THE CHAPERONE ACTIVITY OF THE VESICULAR STOMATITIS VIRUS PHOSPHOPROTEIN

Cédric Leyrat[1], Filip Yabukarski[1], Nicolas Tarbouriech, Euripedes de Almeida Ribeiro Jr[1], Francine Gérard[1], Rob W. H. Ruigrok[1], and Marc Jamin[1*]

[1]Unit of Virus Host Cell Interactions, UMI 3265 UJF-EMBL-CNRS
6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

Running title: Structure of the N^0-P complex

* **Corresponding author:** Unit of Virus Host Cell Interactions (UVHCI), UMI 3265 UJF-EMBL-CNRS, 6, rue Jules Horowitz, B.P. 181, 38042 Grenoble Cedex 9, France
E-mail : jamin@embl.fr, Phone: + 33 4 76 20 94 62, Fax : + 33 4 76 20 94 00

ABSTRACT

The genomic RNA of vesicular stomatitis Virus is encapsidated by the nucleocapsid protein (N) and forms an N-RNA complex that constitutes the template for transcription and replication. During viral replication, the phosphoprotein (P) forms a complex with nascent N protein, named the N^0-P complex, that prevents the oligomerization of N and the encapsidation of cellular RNAs. The N^0-P complex thus mediates the specific encapsidation of viral RNA. It relies on the interaction between N and the N-terminal moiety of P. On the basis of the N-RNA structure, we developed a strategy for reconstructing a N^0-P complex from purified components. N was produced as mutant deleted from its first 21 amino-acids ($N_{\Delta 21}$) and fused to a maltose binding protein (MBP) whereas P was reduced to a peptide corresponding to the 60 first amino-acids (P_{NTD}). We show that P_{NTD} is able to bind and to convert $N_{\Delta 21}$-RNA oligomers into a RNA free complex. We characterize the structure of the $N_{\Delta 21}{}^0$-P_{NTD} complex in solution using small angle X-ray scattering (SAXS) and show that it corresponds to a monomeric N protein bound to P_{NTD}. The complex was crystallized and we report its structure at 3.0 Å resolution, which reveals how P_{NTD} binding prevents wild type N oligomerization and RNA encapsidation by binding on a site of N that overlaps with the RNA binding site but also with the N-N binding site. Furthermore, using site-directed mutagenesis, molecular dynamics and protein-protein docking, we show that the 10 first amino-acids of P are engaged in an interaction with the nucleoprotein C-loop, thus contributing to prevent $N_{\Delta 21}$ oligomerization and stabilizing the N^0-P complex. These results provide the first structural insights into the mechanism of N^0-P formation and its function in the viral cycle.

INTRODUCTION

The *Rhabdoviridae* is a family of viruses whose genome is made of a non-segmented single stranded negative-sense RNA molecule, and which includes human (rabies virus - RV), animal (vesicular stomatitis virus - VSV) and plant pathogens. In the *Rhabdoviridae*, as well as in the other single stranded negative-sense RNA viruses, the viral genome is encapsidated by the nucleoprotein (N), forming a long helical N-RNA complex (Fields, Knipe et al. 1996) that packs the RNA in the viral particle, correctly displays the RNA genome for both transcription and replication by the viral polymerase complex (Arnheiter, Davis et al. 1985) and that certainly protects the RNA genome against nucleases and the innate immune system. In VSV, RNA transcription and replication requires the presence of three viral proteins: the nucleoprotein (N), the phosphoprotein (P) and the

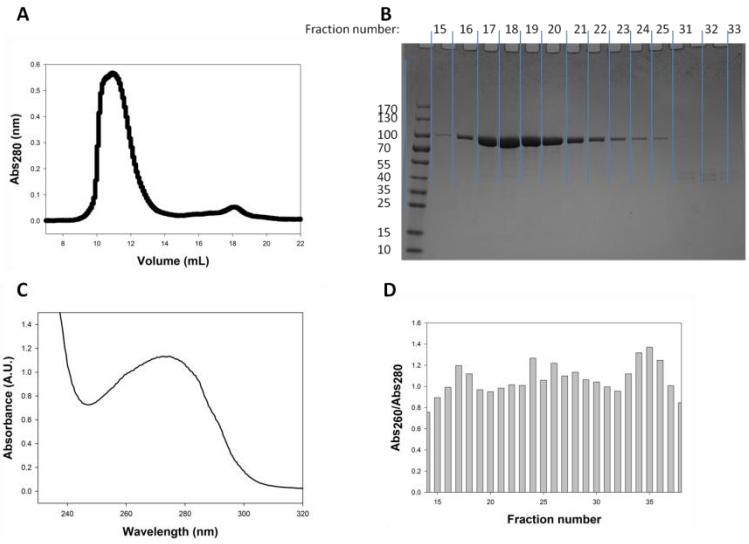


Figure 1: Biochemical characterization of the $N_{\Delta 21}$-RNA-MBP complexes. A. Elution profile of the MBP-$N_{\Delta 21}$-RNA complex on a Superdex 200 column. The experiment was performed at 4 °C. **B.** SDS-PAGE loaded with 5 µl from representative fractions collected from the elution profile represented in A. Fractions (15-25 correspond to the main peak which elutes between 10 and 14 ml, while fractions 31, 32 and 33 correspond to the second minor peak at 18 ml). Based on their apparent molecular weight on the gel, the main peak corresponds to $N_{\Delta 21}$ and the minor peak corresponds to free, self-cleaved $N_{\Delta 21}{}^0$ and MBP. **C.** UV absorption spectrum of MBP-$N_{\Delta 21}$-RNA showing the presence of RNA in the sample. D. Abs260/Abs280 *ratio* of the fractions collected from the elution profile in A, indicating that all fractions contain RNA.

large subunit of the RNA-dependent RNA polymerase (L). The phosphoprotein (P) is involved in different stages of the viral replication process (Emerson and Yu 1975; Masters and Banerjee 1988). Firstly, it associates with the L protein to form an active RNA polymerase that catalyzes both transcription and replication of the viral genome. In this complex, L carries out the RNA synthesis and the 5' and 3' RNA processing required for the production of messenger RNAs, whereas P acts as an essential cofactor (Emerson and Yu 1975) by attaching L onto its template (Mellon and Emerson 1978; Emerson and Schubert 1987). Indeed, by itself, L is unable to bind to the N-RNA complex (Mellon and Emerson 1978), whereas P possesses binding sites for both L and the N-RNA complex and thus ensures the correct positioning of the polymerase onto its template. Secondly, P is essential for encapsidating newly synthesized RNA genomes (Peluso 1988; Peluso and Moyer 1988; Howard and Wertz 1989). Although transcription of the VSV genome can be reconstituted from purified L, P and N-RNA template (Emerson and Yu 1975; De and Banerjee 1984), replication requires the continuous supply of nascent N protein (Hill, Marnell et al. 1981; Peluso and Moyer 1983; Patton, Davis et al. 1984; Davis, Arnheiter et al. 1986). To prevent from binding non-specifically to cellular RNAs, P binds to nascent N molecules and forms a N^0-P complex (the superscript ^{0} denotes the absence of RNA) that maintains N in a soluble form, available for the encapsidation of newly synthesized RNA genomes (Masters and Banerjee 1988; Peluso and Moyer 1988; Howard and Wertz 1989; Mavrakis, Iseni et al. 2003; Mavrakis, Mehouas et al. 2006; Chen, Ogino et al. 2007). Upon formation of viral nucleocapsids, the N^0-P complex dissociates. N is transferred to the viral RNA, while P is released (Peluso and Moyer 1988). In both VSV and RV P, the N-terminal region of the protein (P_{NTD}) was identified as the major site of interaction with N^0. Residues 4 to 40 in RV and residues 11 to 30 in VSV are sufficient for maintaining N in a soluble form (Mavrakis, Mehouas et al. 2006; Chen, Ogino et al. 2007) . The crystal structure of circular N-RNA complexes from both VSV and RV revealed that the N protein contains a N-terminal domain (NTD) and a C-terminal (CTD) domain separated by a positively charged groove which binds the RNA (Albertini, Wernimont et al. 2006; Green, Zhang et al. 2006). NTD and CTD act like jaws that close around and completely enwrap the RNA, forming multiple salt bridges with the phosphate backbone of the RNA. Because P_{NTD} is rich in negatively charged residues, it was hypothesized that P_{NTD} binds in the RNA-binding cavity of N, thereby preventing the non-specific attachment of N^0 to cellular RNA (Hudson, Condra et al. 1986; Mavrakis, Mehouas et al. 2006). N also contains N- and C- terminal subdomains that stabilize the assembly of multiple N molecules by exchanging with neighboring molecules (Green, Zhang et al. 2006). Until now, the difficulties encountered in preparing the N^0-P complex by co-expressing N and P precluded its detailed structural characterization (Mavrakis, Iseni et al. 2003). Here, we report a new strategy developed for producing a N^0-P complex from.

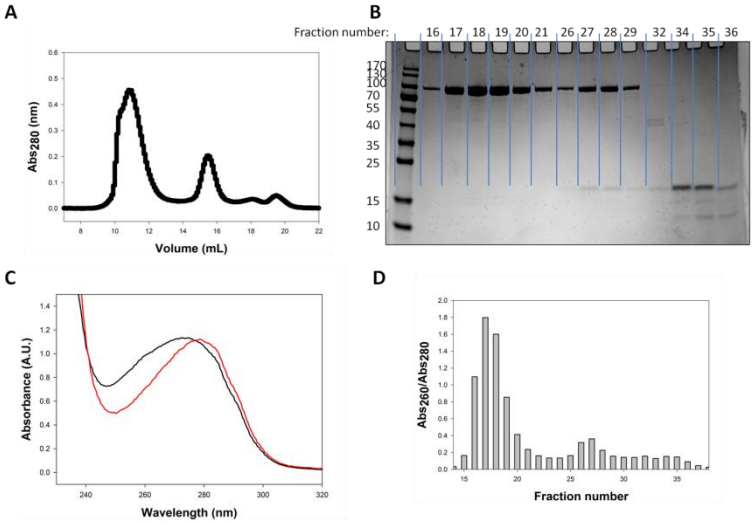


Figure 2: Formation of the MBP-$N_{\Delta 21}{}^{0}$-P_{NTD} complex. A. Elution profile of MBP-$N_{\Delta 21}$-RNA after 1 hour of incubation with a slight excess of P_{NTD} on a Superdex 200 column. The experiment was performed at 4°C. **B.** SDS-PAGE loaded with 5 µl from representative fractions collected from the elution profile represented in A. Fractions (16-21 correspond to the main peak which elutes between 10 and 14 ml, fractions 26-29 correspond to the MBP-$N_{\Delta 21}{}^{0}$-P_{NTD} complex and show two bands around 100 kDa and 15 kDa, respectively, fraction 32 corresponds to the minor peak at 18 ml that is also observed in the elution profile from Figure 1A, and fractions 34-36 contain free P_{NTD}). **C.** UV absorption spectrum of $N_{\Delta 21}$°-P_{NTD}-MBP (red) showing the absence of RNA in the sample. The UV spectrum of MBP-$N_{\Delta 21}$-RNA is superimposed in black for comparison. The spectra were normalized for presentation purpose. D. Abs260/Abs280 ratio of the fractions collected from the elution profile in A, indicating an increase in RNA content of the first fractions compared to Figure 1D, and the absence of RNA from the fractions containing MBP-$N_{\Delta 21}$°-P_{NTD} complex.

individually purified components and the structure at low and high resolution of this complex

RESULTS

Strategy for producing the N^0-P complex

Addition of a synthetic peptide corresponding to the 60 first amino acids of P (P_{NTD}) to recombinant circular N-RNA complexes was unable to displace the RNA molecule from a purified N-RNA complex. On the basis of the N-RNA structure, we reasoned that deletion of the N-terminal sub-domain may destabilize the N-RNA complex sufficiently for allowing P_{NTD} to displace the RNA molecule. A mutated form of N lacking the 21 first N-terminal residues ($N_{\square 21}$) was produced in *E. coli* with a N-terminal maltose binding protein (MBP) tag as a 100 kDa protein (Figure 1 B (SDS PAGE)). The MBP-$N_{\Delta 21}$ protein binds RNA as evidenced by its absorbance spectrum and A260/A280 *ratio* (Figure 1 C&D) and self-assembles as shown by its elution in the exclusion volume of the size exclusion chromatography column (Figure 1 A). However, the MBP-$N_{\Delta 21}$-RNA complex was less stable than the native N-RNA complex, and incubation in the presence of P_{NTD} resulted in the dissociation of the oligomers and in the formation of a MBP-$N_{\Delta 21}{}^{0}$-P_{NTD} complex without RNA (Figure 2). The complex elutes in the separation range of the SEC, corresponding to a hydrodynamic radius of 4.5 nm. SDS-PAGE revealed the presence of both proteinsin this complex (Figure 2 B), and the absorbance spectrum shows that in contains no RNA (Figure 2 C). Also, the A260/A280 *ratio* of the material eluting in the exclusion volume increases, suggesting that the released RNA now elutes in the exclusion volume (Figure 1 D and Figure 2 D). After cleavage of the MBP tag with the TEV protease, the complex could be purified by Ni^{2+} chelate affinity chromatography by using the His-tag present on P_{NTD} and by SEC. The analysis of the $N_{\Delta 21}{}^{0}$-P_{NTD} complex by SEC-MALLS-RI (Figure 3) showed that the preparation is monodisperse and yielded an average molecular mass (MM) of 55 kDa ± 5kDa that clearly indicates a 1:1 stoichiometry in agreement with the calculated MM of 53,399 Da.

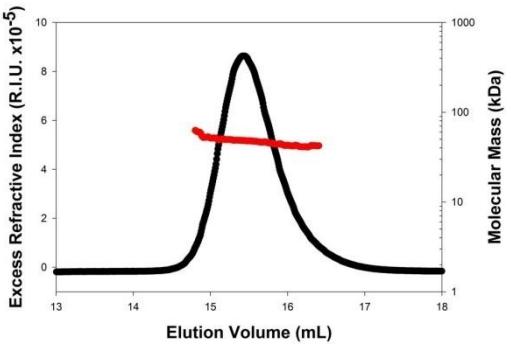


Figure 3: SEC-MALLS elution profile of the $N_{\Delta 21}$°-P_{NTD} complex. The black line shows the elution profile as monitored by refractometry. The red circles show the molecular mass calculated from light scattering and refractometry data.

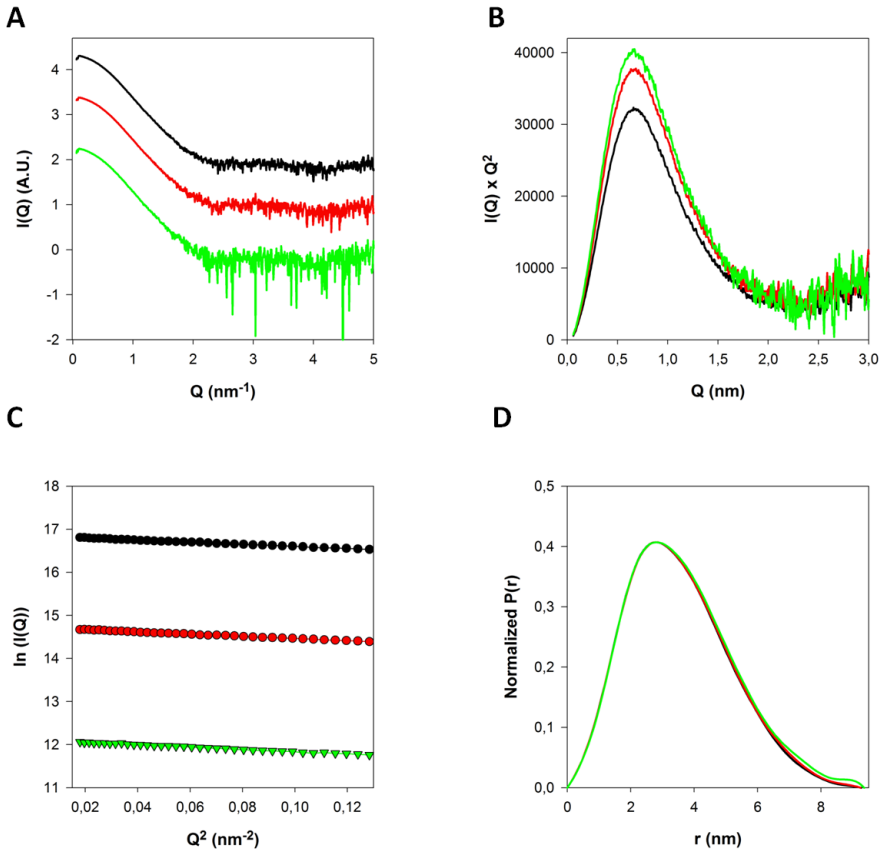


Figure 4: Small Angle X-ray Scattering experiments. A. SAXS data were recorded for Q values in the range 0.05 nm^{-1}< Q < 5nm^{-1}. The curves obtained for 3 protein concentrations (2.7mg/ml, 5.3mg/ml, and 8mg/ml) are represented in green, red, and black lines, respectively. **B.** Kratky plot at three concentrations. The typical bell-shape of the curve indicates that the protein is mostly folded. **C.** Guinier analysis was performed for each three concentrations and confirms the absence of absence of concentration dependent aggregation in the sample. **D.** Normalized pair distribution function P(r) at three concentrations.

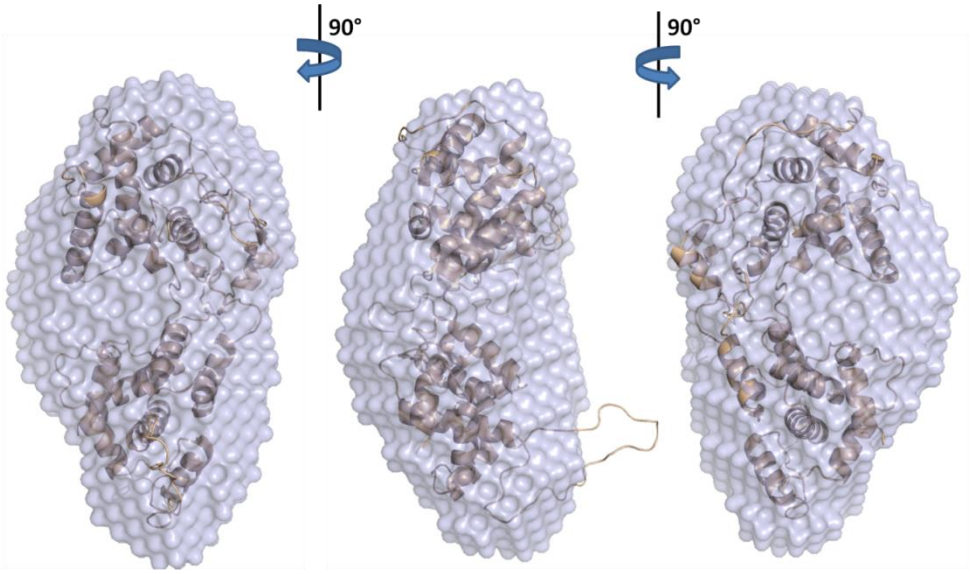


Figure 5: *Ab initio* models – DAMMIF models. The average model from 20 independent runs is represented as a surface and three views shifted by 90° are shown. The structure of a nucleoprotein protomer from the N-RNA crystal structure is superimposed onto the density.

Low resolution structure of the N^0-P complex in solution from SAXS

The SAXS profile of the $N_{\Delta 21}{}^0$-P_{NTD} complex was recorded for scattering vectors in the range 0.05 nm^{-1} < Q < 4.0 nm^{-1} at three different protein concentrations (Figure 4 A). No significant dependence of the radius of gyration (R_g) on protein concentration was detected, revealing the absence of aggregation or intermolecular interactions in the

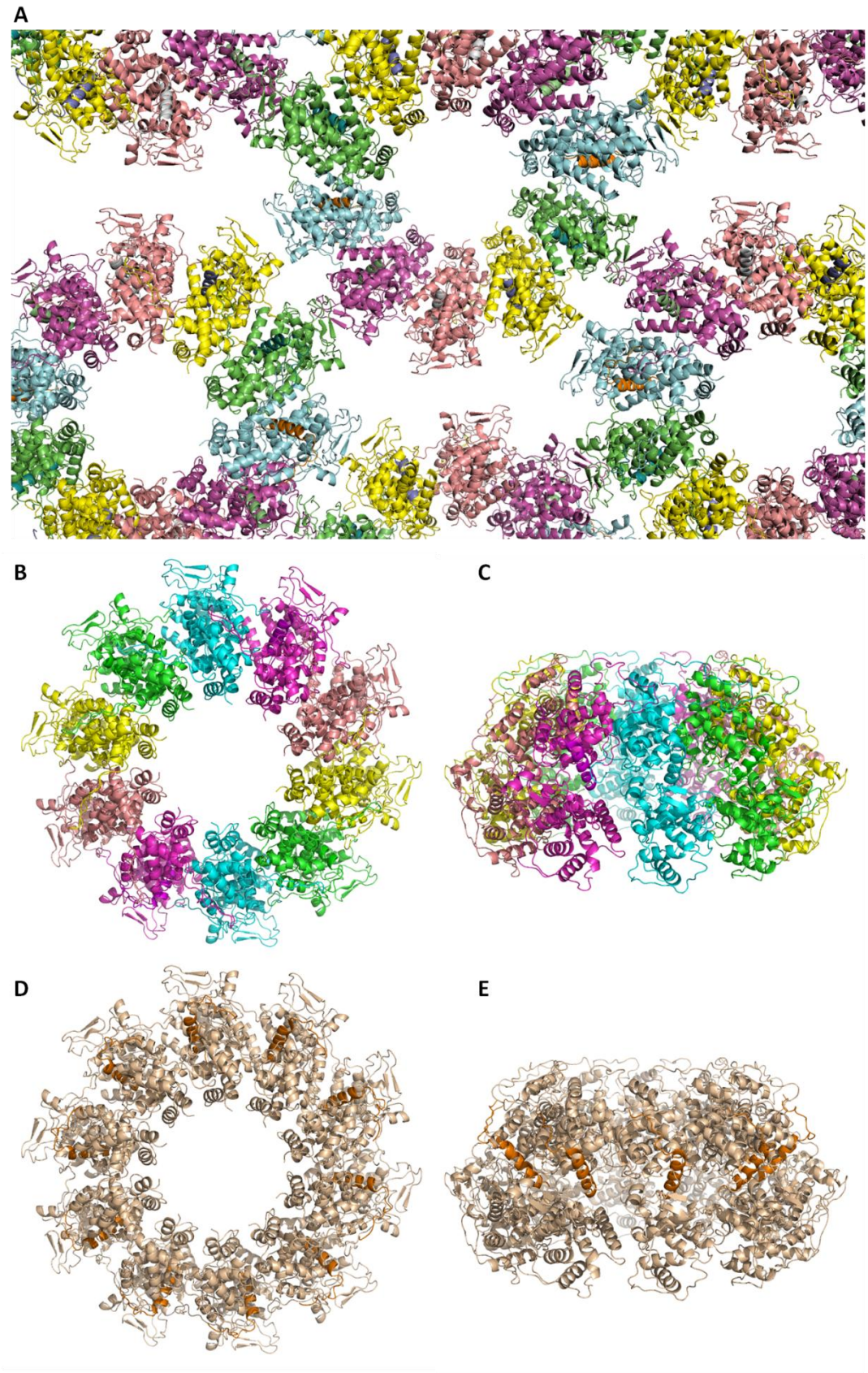


Figure 6: Crystal packing in the structure of the $N_{\Delta 21}$°-P_{NTD} complex. A. Crystallographic arrangement of the $N_{\Delta 21}{}^{0}$-P_{NTD} sub-units. The asymmetric unit is composed of 5 sub-units each one comprising one monomer of $N_{\Delta 21}{}^{0}$ and one monomer of P_{NTD}. The neighboring asymmetric units were generated from the crystal lattice symmetry using PYMOL. **B and C.** Cartoon representation of the $N_{\Delta 21}{}^{0}$-P_{NTD} rings in two views shifted by 90° highlighting the presence of P_{NTD} (orange) between each N protomer. **D and E.** Cartoon representation of the $N_{\Delta 21}{}^{0}$-P_{NTD} rings in two views shifted by 90° highlighting the presence of 10 N protomers *per* ring.

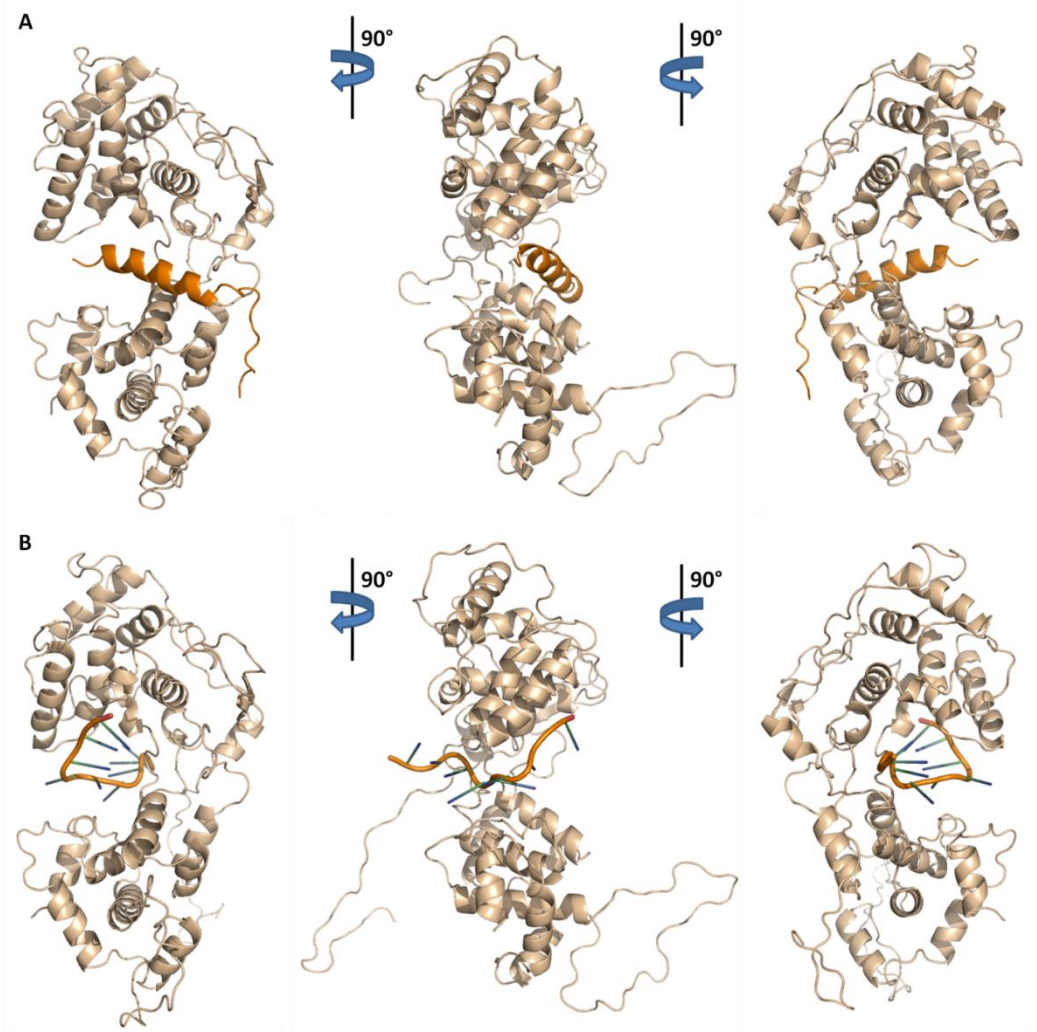


Figure 7: Comparison between the N-RNA and the $N_{\Delta 21}{}^{0}$-P_{NTD} structures. A. Three views of the $N_{\Delta 21}{}^{0}$-P_{NTD} protomer shifted by 90° showing the localization of P_{NTD} onto the nucleoprotein. **B.** Three views of the N-RNA protomer bound to 9 nucleotides and shifted by 90°.

concentration range used (2.7 - 8 mg/ml). The measured R_g value of 2.7 ± 0.1 nm (Figure 4 C), and the MM of 65 ± 15 kDa derived from the scattering intensity at zero angle, I_0, were in agreement with SEC-MALLS-RI results. Additionally, the distance distribution function, P(r) (Figure 4 D), and the Kratky plot (Figure 4 B) are typical for globular, folded proteins. The low resolution *ab initio* shape of the complex reconstructed with the program DAMMIF (Franke and Svergun 2009) displayed a typical bean shape reminiscent of a nucleoprotein protomer. However, the RNA binding cavity of N is filled and the presence of a small lobe on the side of the protein could indicate the location of the C-loop (Figure 5).

Crystallographic structure of the N°P complex

The $N_{\Delta 21}$°-P_{NTD} complex was crystallized in space group $P2_12_12$ and diffraction data were collected up to a resolution of 3.0 Å. The structure was solved by molecular replacement using a N protomer from the N-RNA crystal structure (pdb id: 2GIC). A clear density map was obtained in which the atomic model was built and refined. In the crystallographic conditions, $N_{\Delta 21}{}^{0}$-P_{NTD} crystallized as a 10-membered ring similar to that formed by the wild-type N-RNA complex (Green, Zhang et al. 2006) (Figure 6 A, B&C). However, RNA was absent from the structure and an additional density corresponding to residues 5 to 33 of P_{NTD} was visible in a region close to the RNA-binding cavity (Figure 6 D&E, Figure 7 A). The residues 17 to 31 of P_{NTD} form an amphipathic α-helix that lies on the 3' side of the RNA-binding cavity, rendering impossible the interaction of N with RNA (Figure 7 A&B, Figure8 C&D). This α-helix makes extensive contacts within the RNA binding cavity of N (Figure 9). The binding is stabilized by a network of hydrophobic interactions involving residues

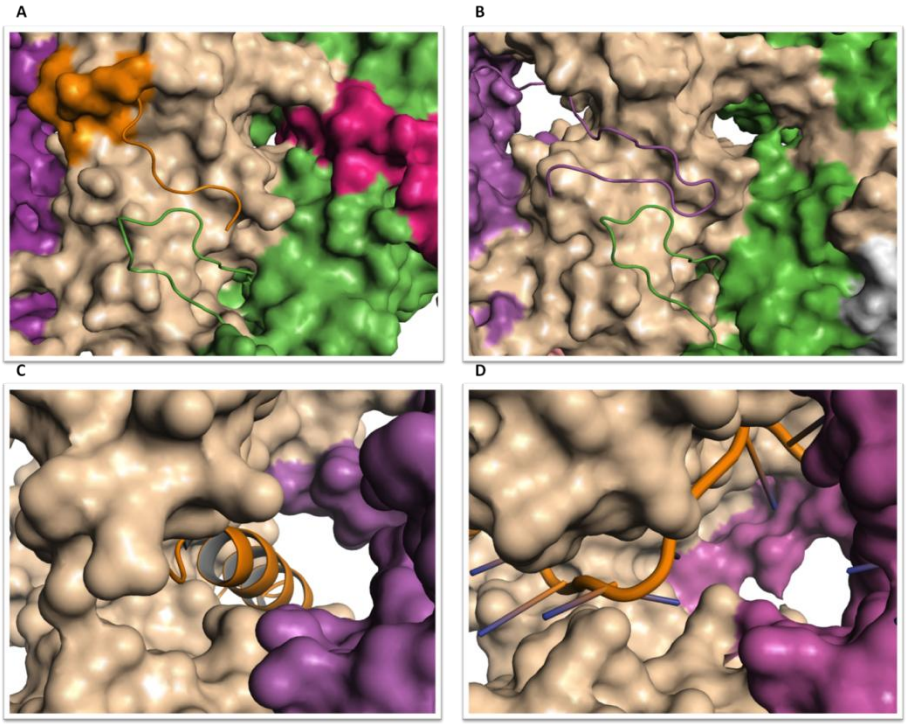


Figure 8: Conflicting localization of P_{NTD} in the $N_{\Delta21}{}^{0}$-P_{NTD} structure with the localization of the RNA and the N-terminal arm of N in the N-RNA structure. A and B. Close-ups of the interaction between the C-loop of N and the N-terminal amino-acids of P_{NTD} or the N-terminal subdomain of N in the $N_{\Delta21}{}^{0}$-P_{NTD} and in the N-RNA structures, respectively. **C and D.** Close-ups of the interaction between the RNA binding cavity of N and the α-helix of P_{NTD} or an RNA molecule in the $N_{\Delta21}{}^{0}$-P_{NTD} and in the N-RNA structures, respectively.

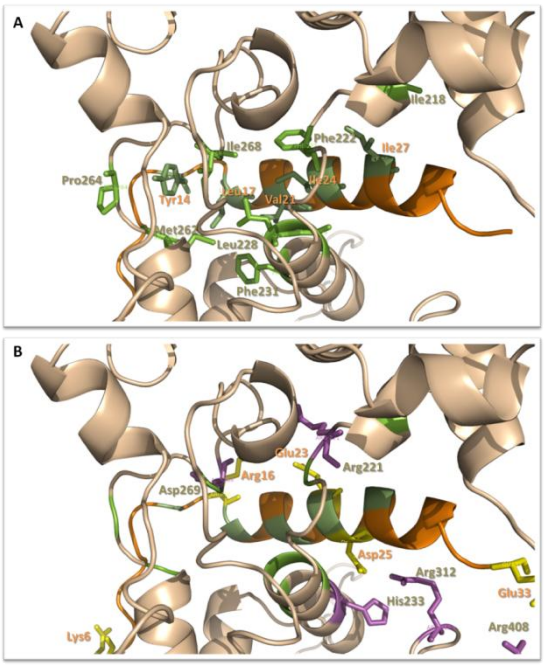


Figure 9: Hydrophobic and ionic interactions involved in stabilizing the $N_{\Delta21}{}^{0}$-P_{NTD} complex. A. Main stabilizing hydrophobic interactions. Interacting residues from P_{NTD} and from N are shown as sticks and colored in green. The names and numbers of interacting residues from $N_{\Delta21}{}^{0}$ and P_{NTD} are indicated in wheat and in orange, respectively. **B.** Main stabilizing ionic interactions. Interacting residues from P_{NTD} and from N are shown as sticks and colored in yellow and purple, respectively. Again, the names and numbers of interacting residues from $N_{\Delta21}{}^{0}$ and P_{NTD} are indicated in wheat and in orange, respectively.

Leu17, Val21, Ile24, Ile27 (i+3 or i+4 spacing) of the P_{NTD} α-helix and residues Ile218, Phe222, Leu228, Phe231, Met262, Ile270 of N (Figure 9A). The hydrophobic cluster extends beyond the α-helix of P_{NTD} and also includes Val7, Tyr10, and Tyr14 that interact with Val259, Met262, Pro264 and Ile268 of N (Figure 9A). Six salt bridges provide additional stability to the interface (Figure 9B). Two of these ionic interactions involve the negatively charged Asp256 and Asp 269 from N and Lys6 and Arg16 of P, respectively, while the four remaining salt bridges are formed by Arg221 and Glu23, His233, Arg312 and Asp25, and Arg408 and Glu33 (Figure 9B). The structure is further stabilized by 14 intermolecular hydrogen bonds. In particular, the hydroxyl group of Tyr14 is hydrogen-bonded to the backbone carbonyl group of Leu263 and to the backbone amide group of Gln266, the carboxyl group Asp18 with the side-chain hydroxyl group and backbone amide group of Thr242, and the carbonyl group of Gln30 with the side-chain hydroxyl group of Tyr152 and backbone amide group of Arg143. Arg143 and Arg408 of N, which are coordinating sugar phosphates in the N-RNA complex, are also directly implicated in the interaction with P_{NTD}, whereas several other positively charged residues in N, that are involved in RNA binding, are part of the interface but make no specific contacts with P_{NTD}. Interestingly,

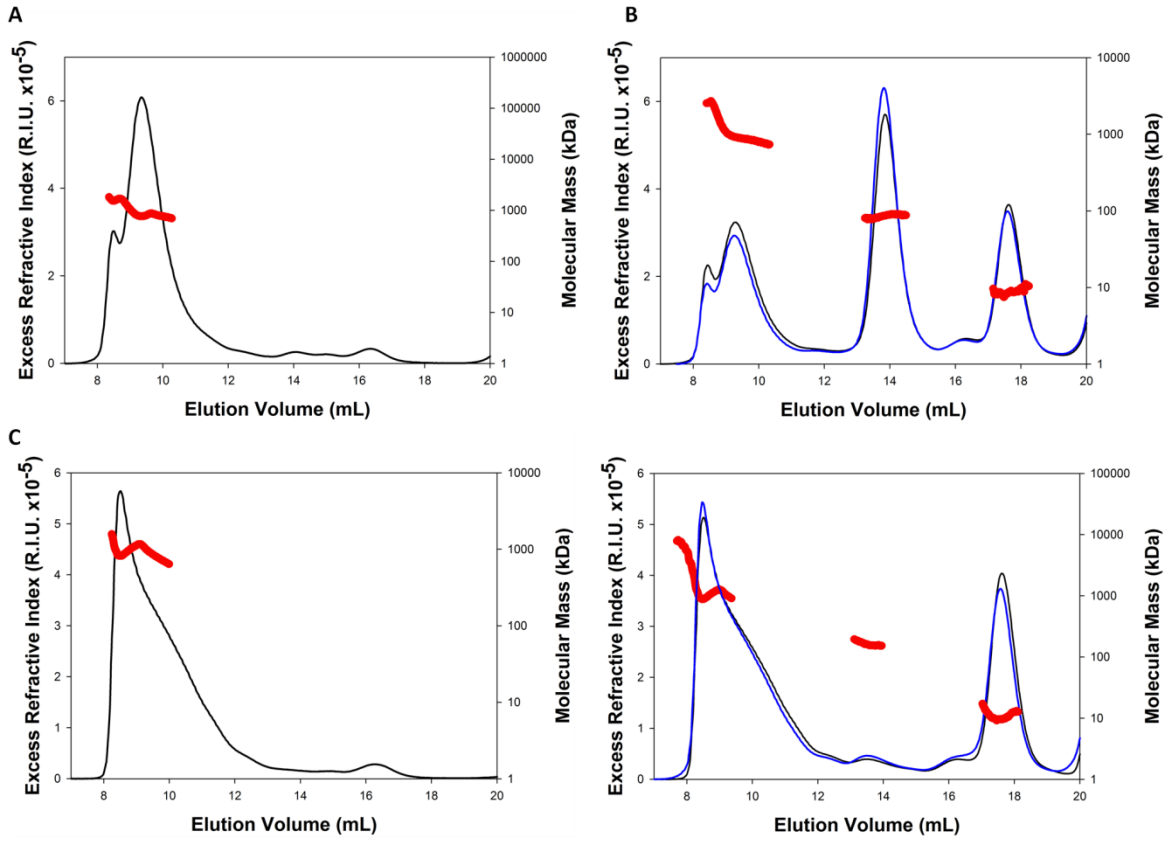


Figure 10: Comparison between SEC-MALLS elution profiles of MBP-$N_{\Delta 21}$ and MBP-$N_{\Delta 21}^{344\text{-SPSGS-}348}$ before and after incubation with P_{NTD}. A. SEC-MALLS elution profile of MBP-$N_{\Delta 21}$-RNA. The black line shows the elution profile as monitored by refractometry. The red circles show the molecular mass calculated from light scattering and refractometry data. **B.** SEC-MALLS-RI elution profile of MBP-$N_{\Delta 21}$-RNA incubated with a large excess of P_{NTD} for one hour (black line) or overnight (blue line). The black and blue lines show the elution profiles as monitored by refractometry. The red circles show the molecular mass calculated from light scattering and refractometry data from the overnight incubated sample. **C.** SEC-MALLS-RI elution profile of MBP-$N_{\Delta 21}^{344\text{-SPSGS-}348}$. The black line shows the elution profile as monitored by refractometry. The red circles show the molecular mass calculated from light scattering and refractometry data. **D.** SEC-MALLS-RI elution profile of MBP-$N_{\Delta 21}^{344\text{-SPSGS-}348}$ incubated with a large excess of P_{NTD} for one hour (black line) or overnight (blue line). The black and blue lines show the elution profiles as monitored by refractometry. The red circles show the molecular mass calculated from light scattering and refractometry data from the overnight incubated sample.

residues 6 to 9 of P_{NTD} are contacting residues 346 to 350 of the C-loop from the neighboring N protomer (Figure 8 A), occupying the surface pocket emptied by the deletion of the N-terminal arm. Moreover, a structural alignment of the $N_{\Delta 21}^{0}$-P_{NTD} with the N-RNA complex (data not shown) reveals major structural clashes between residues 6 to 21 of P_{NTD} and the N-terminal subdomain of N that was deleted in our construct, suggesting how P_{NTD} could prevent the self-assembly of wild type N (Figure 8A&B). However, the ability of $N_{\Delta 21}$ to form a monomeric $N_{\Delta 21}^{0}$-P_{NTD} complex at neutral pH indicates that an additional mechanism exists that stabilizes monomeric N.

Site-directed mutagenesis and the involvement of the C-loop in binding

We engineered a $N_{\Delta 21}$ mutant bearing the substitution ^{344}LAQQF^{348}→^{344}SPSGS^{348} along with the deletion of the N-terminal arm. This "double" mutant was originally designed to disrupt all the specific N-N contacts leading to nucleoprotein polymerization (Luo, Green et al. 2007). In particular, the substitution of Ala345 to a proline residue destabilizes the backbone conformation adopted by the C-loop in the N-RNA complex, and Leu344 and Phe348, which are involved in hydrophobic interactions, are mutated to hydrophilic serine residues. The MBP-$N_{\Delta 21,344\text{-SPSGS-}348}$ formed a multimeric complex that elutes in the exclusion volume of the SEC column but contains no RNA (Figure 10c). We found that the MBP-$N_{\Delta 21,344\text{-SPSGS-}348}$ was significantly impaired in its ability to bind P_{NTD}, that occurred only in the presence of a large excess of P_{NTD} (Figure 10B&D). The very low affinity for P_{NTD} observed for MBP-$N_{\Delta 21,344\text{-SPSGS-}348}$ compared to MBP-$N_{\Delta 21}$ suggested the involvement of the nucleoprotein C-loop in

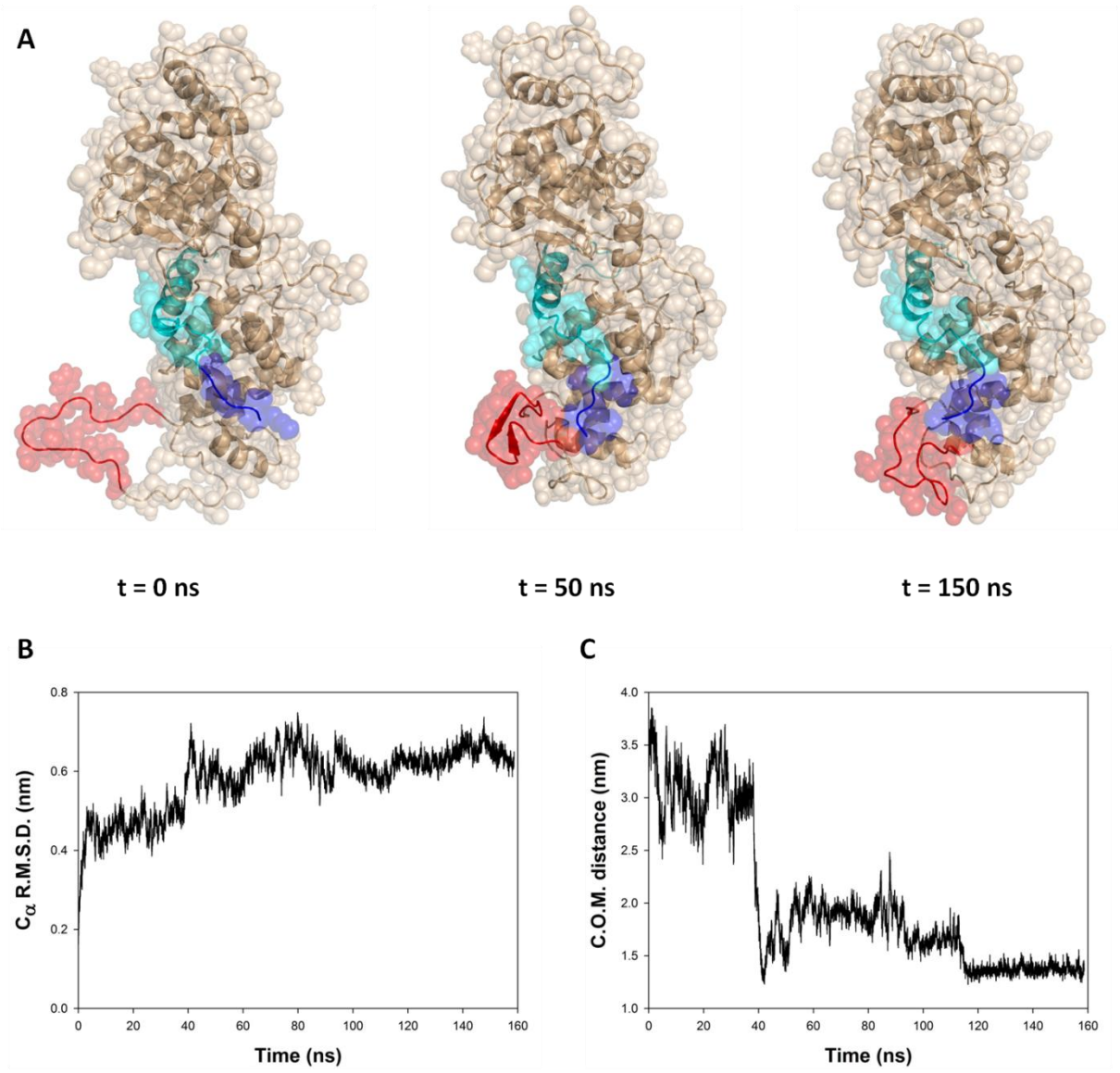


Figure 11: Molecular Dynamics Simulation of the $N_{\Delta 21}{}^{0}$-P_{NTD} complex. A. Three snapshots from the simulation taken at t = 0 ns , t = 50 ns and t = 150 ns. Residues 342-358 of N are colored in red, and residues 6-10 of P_{NTD} are in blue. **B.** Cα r.m.s.d. of the $N_{\Delta 21}{}^{0}$-P_{NTD} complex. The r.m.s.d. increases quickly during the first 10 ns of the simulation and shows a significant conformational change around t = 40 ns. The value tends to stabilize around 0.6 nm in the latter part of the trajectory. **C.** Distance between the center of mass of the residues 6 to 10 of P_{NTD} and the centre of mass of residues 342 to 358 of N.

binding P_{NTD}. In order to gain insights into this interaction, we turned to molecular modeling.

Computational model from protein docking and molecular dynamics simulation

First, we ran a 160 ns molecular dynamics simulation (MDS) in explicit solvent using a sub-unit from the crystal structure, composed of $N_{\Delta 21}$ and residues 6 to 40 of P_{NTD} (Figure 11). The simulation showed large motions of the C-loop during the first 40 ns, followed by a spontaneous binding event between residues 6 to 9 of P_{NTD} and a cluster of residues from the C-loop at t = 40 ns, which is stabilized at t = 120 ns. In order to model more accurately this interaction, we used flexible protein – protein docking. A previous NMR and SAXS study revealed several features of the conformational behavior of P_{NTD} in solution (Leyrat et al., 2010), and notably the presence of two segments containing partially populated α-helices. These helices located between residues 1 to 12 and between residues 25 to 38 are locally encoded in the amino acid sequence. We assumed that the N-terminal helix is transiently present in solution and therefore modeled residues 1-13 as an α-helix and

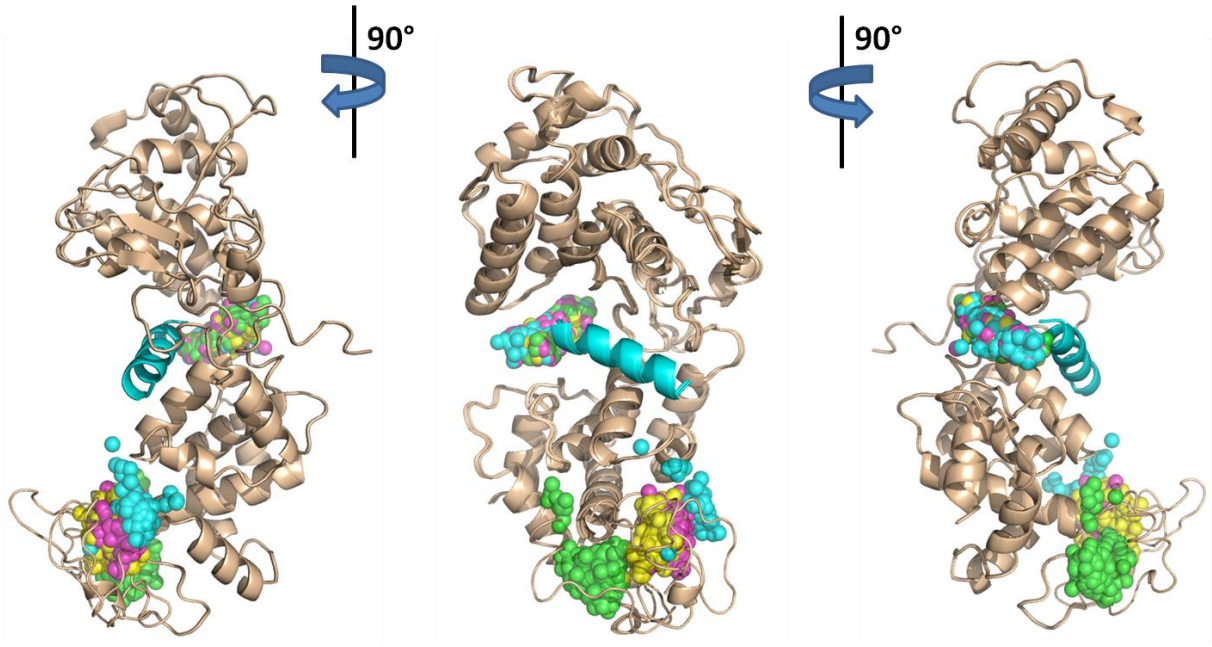


Figure 12: Computational docking of the fragment 1 to 13 of P_{NTD} on the nucleoprotein. Three views shifted by 90°C are shown. The 100 best scored solutions of 4 representative docking runs are shown (yellow, cyan, magenta, green) each one with a different conformation of residues 343-363 of the nucleoprotein C-loop. The localization of P_{NTD} residues 1 to 13 onto N is indicated as a sphere placed at its center of mass, which suggests that P_{NTD} residues 1 to 13 bind the C-loop.

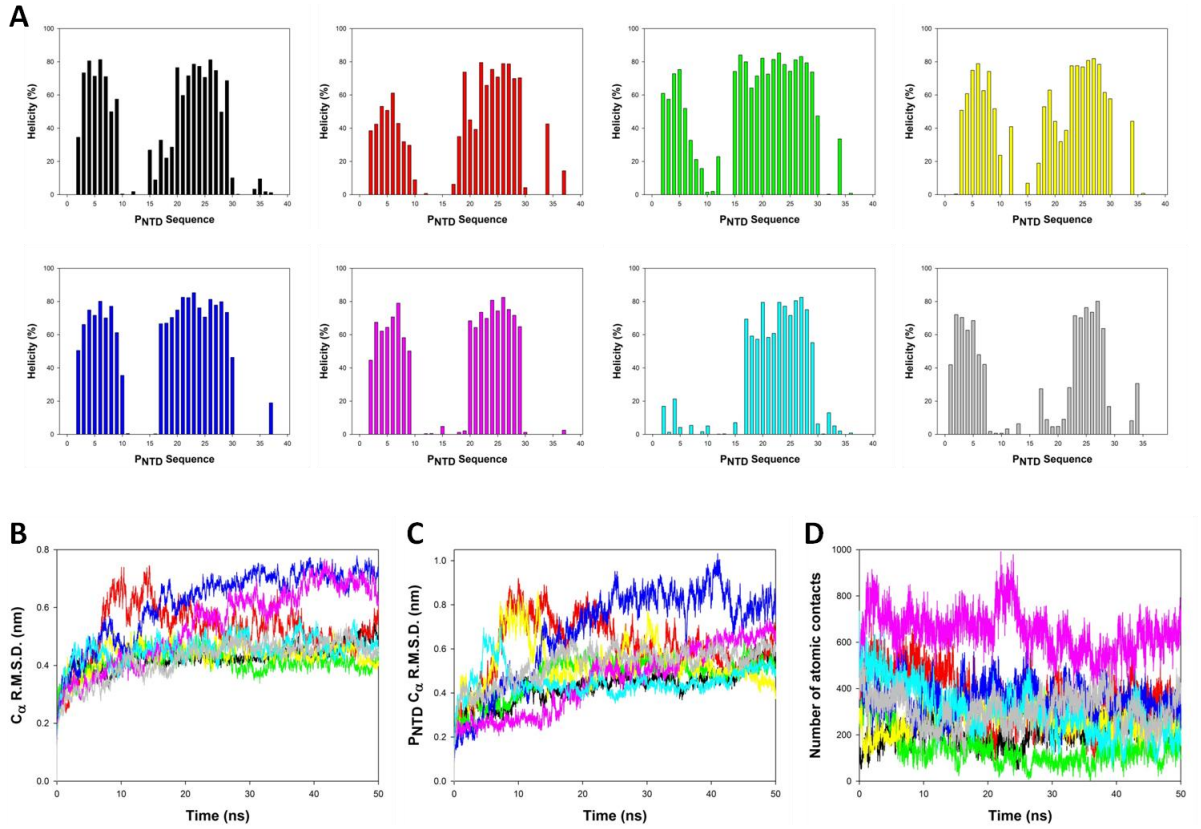


Figure 13: Molecular dynamics simulation of the 8 best scored $N_{\Delta 21}{}^{0}$-P_{NTD} decoys. A. Helicity histograms showing the percentage of simulation time (based on 50 ns of simulation) each residue of P_{NTD} adopts an α-helical conformation for the 8 best scored decoys. **B and C.** Cα r.m.s.d. of the $N_{\Delta 21}{}^{0}$-P_{NTD} decoys (B) and P_{NTD} alone (C). The r.m.s.d. increases quickly during the first 10 ns of the simulation and usually tends to stabilize over simulation time. **D.** Number of intermolecular contacts (<5Å) between P_{NTD} residues 1 to 13 and the C-loop of N, which allows to monitor unbinding events.

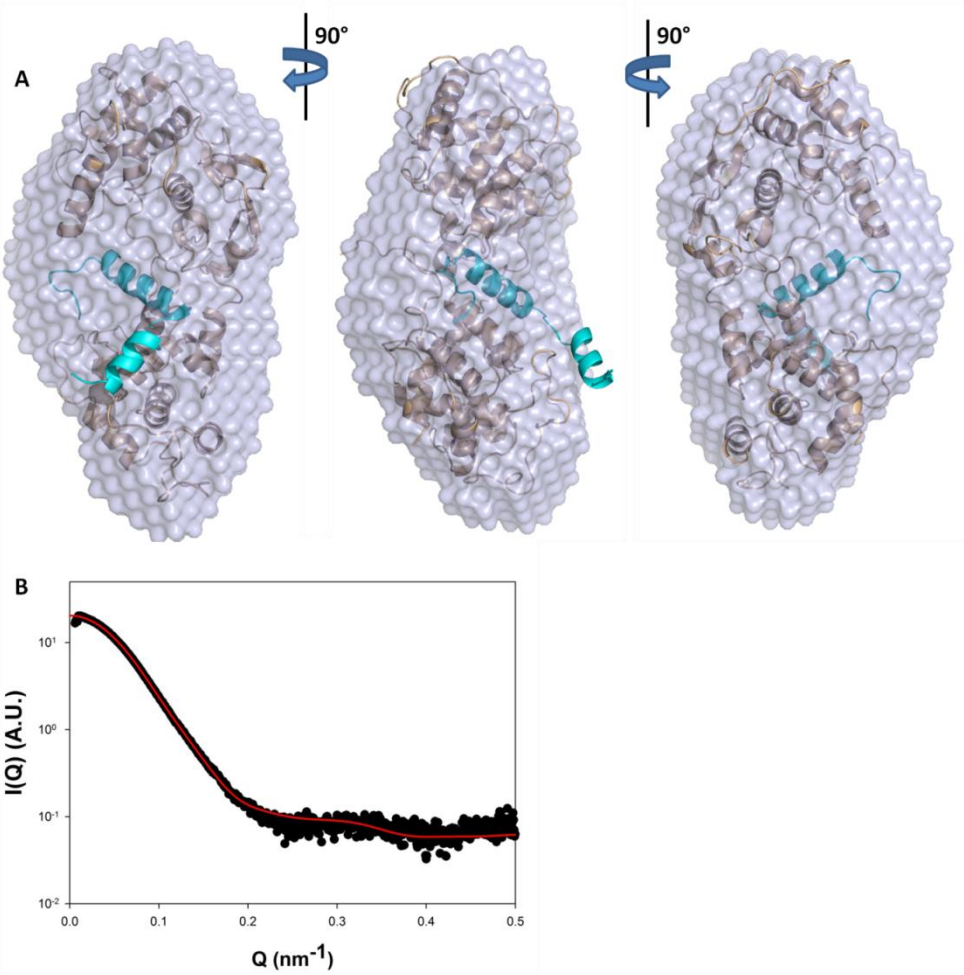


Figure 14: Fitting of SAXS data. A. The low resolution electron density retrieved from SAXS data is superimposed into the best scored decoy from docking and MD simulations. Three views shifted by 90° are shown. **B.** Fitting of the SAXS profile using the best model obtained from the docking procedure.

docked this fragment onto 30 conformers of $N_{\Delta 21}$ containing the residues 17 - 31 of P_{NTD} (helical part) but exhibiting a different C-loop conformation, in order to account for the conformational disorder of the C-loop in the monomeric state. From this calculation, we observe two different types of clusters of solutions which are located in the area of the C-loop or in the center of the RNA binding cavity, respectively. As the second cluster of solutions doesn't allow for a physical link between the two P_{NTD} fragments, the only remaining physically relevant clusters of molecular decoys observed within the 100 best scores of multiple docking runs involved binding of the 1-13 fragment of P_{NTD} to the C-loop of N (Figure 12). We built residues 14-16 and 32-40 of P_{NTD} in the 8 scored decoys in order to obtain models of the complex including the first 40 residues of P_{NTD} and we refined their structure by 50 ns MDS. We monitored various simulation parameters, including α-helicity (Figure 13 A), r.m.s.d. (P_{NTD} only or whole structure) (Figure 13 B&C), and intermolecular contacts (Figure 13 D). We used the optimization of theses parameters and the agreement with SAXS data to rank the models. The best model that was obtained from this refinement and scoring procedure, both in terms of agreement with the SAXS profile (Figure 14) and the secondary structure propensity of the isolated P_{NTD} in solution derived from NMR (Figure 13 A), is presented in Figure 15. In this structure, residues 344 to 348 of N extensively interacted with residues 4 to 14 of P_{NTD}, thus explaining the loss of affinity observed for the MBP-$N_{\Delta 21,344\text{-SPSGS-}348}$ mutant. In particular, Leu344 and Phe348 form a small hydrophobic cluster by interacting with the side chains of Leu4 and Arg8, while Arg8 is involved in an ionic interaction with Asp343 (Figure 16). The hydroxyl groups of Ser13 and Tyr14 of P_{NTD} interact with the side chain of

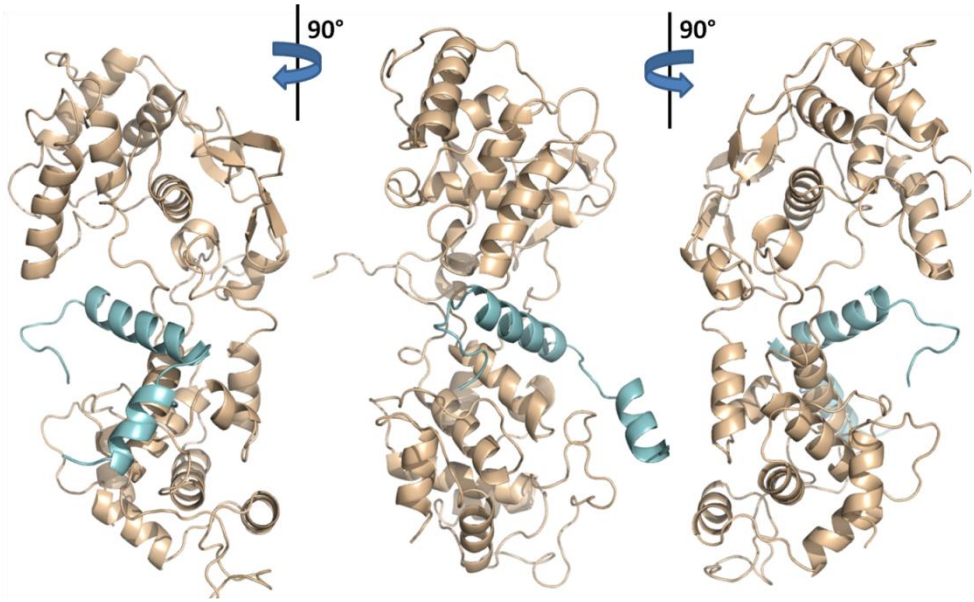


Figure 15: Structure of the computational model of the $N_{\Delta 21}{}^{0}$-P_{NTD} complex. Three views shifted by 90° are shown in cartoon representation. The snapshot is taken from 50 ns of MD simulation. The nucleoprotein is shown in wheat and residues 1 to 40 of P_{NTD} are drawn in cyan.

Gln346, and the backbone carbonyl of Lys12 is hydrogen bonded to the backbone amide group of Ala345. These interactions between the nucleoprotein C-loop and the first 12 amino-acids of P_{NTD} provide an efficient way to sequester the C-loop in a conformation that is inadequate for the self-assembly of N, and therefore contribute to the stability of the monomeric N^{0}-P complex.

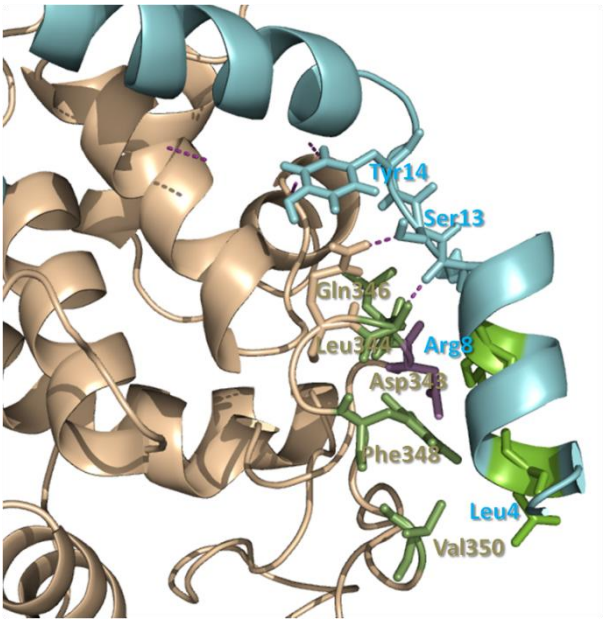


Figure 16: Details of the interaction between N-terminal residues of P_{NTD} and the C-loop of N in the computational model of the $N_{\Delta 21}{}^{0}$-P_{NTD} complex. Interacting residues are shown as sticks. Residues involved in hydrophobic interactions are colored in green. The names and numbers of interacting residues from $N_{\Delta 21}$° and P_{NTD} are indicated in wheat and in orange, respectively. Residues involved in ionic interactions are colored in yellow and purple, as in Figure 9.

DISCUSSION

The structural data reported here provides the first account at the molecular level of the chaperone activity of P. The positioning of P_{NTD} on the N molecule explains how it prevents both non-specific RNA encapsidation and nucleocapsid assembly. P_{NTD} binding site on N overlaps both with the RNA binding site by occupying the 3' side of the central cavity, but also interferes with correct positioning of the N-terminal arm of N that irreversibly locks the oligomeric N-RNA structure. We found that the deletion of the N-terminal arm of N (residues 1 to 21) destabilizes the N-RNA complex and renders the RNA encapsidation reversible in the presence of P_{NTD}. Furthermore, we show that P_{NTD} interacts with the C-loop of N and prevents it from adopting the conformation that is allow the self-assembly of N, and thus stabilizes the monomeric form. A computational model of $N_{\Delta 21}$°-P_{NTD} including residues 1 to 40 of P_{NTD} suggests how the N-terminal part of P_{NTD} interacts with the C-loop of N. The disordered C-loop thus appears as a functional hot spot that controls the fate of the nucleoprotein by interacting with P_{NTD} in the N^{0}-P complex (residues 342 to 351), with N during RNA encapsidation (residues 340 to 354), and with P_{CTD} once N is assembled onto the RNA molecule (residues 354 to 367).

The α-helical structure observed for residues 17 to 31 of P_{NTD} in the crystallographic structure of the $N_{\Delta 21}$°-P_{NTD} provides direct evidence for the binding-induced folding of P_{NTD}, which was

observed to be intrinsically disordered in solution by NMR and to possess partially structured α-helical elements located between residues 1 to 12 and 25 to 38, with the highest population being observed for residues 25 to 31. Residues 17 to 25 possess no helical propensity in solution, which implies that their helical conformation in the crystal is stabilized exclusively by intermolecular interactions. Residues 40 to 60 of P_{NTD} were shown to be dispensable for N^0-P complex formation (Chen, Ogino et al. 2007) and residues 34 to 60 are absent from the structure and presumably disordered in the crystal. However, at least residues 34 to 40 are located in the region of the RNA binding cavity of N as observed in the MD simulations.

MATERIALS & METHODS

Protein Cloning, expression and purification

The vesicular stomatitis virus nucleoprotein gene, and one fragment encompassing residues 22 to 422 were amplified by PCR and cloned into a pET-M40 plasmid (EMBL) using *NcoI* and *XhoI* restriction sites, yielding a N-terminally truncated nucleoprotein bearing a MBP tag at its N-terminus and including a TEV cleavage site. The mutant MBP-$N_{\Delta 21,\ 344\text{-}SPSGS\text{-}348}$ was engineered using the one-step QuickChange® mutagenesis kit (Stratagene). Both constructions were checked using DNA sequencing. The N-terminal region of P corresponding to residues 1 to 60 (P_{NTD}) was cloned into a pET28a plasmid as previously described (Leyrat et al., 2010). Plasmids containing MBP-$N_{\Delta 21}$, MBP-$N_{\Delta 21,\ 344\text{-}SPSGS\text{-}348}$, and P_{NTD} were expressed in *E. coli* Rosetta™ cells. Protein expression was performed overnight at 16 °C after induction with I.P.T.G. at a concentration of 1 mM when culture medium O.D. reached 0.6 A.U. MBP-$N_{\Delta 21}$ and MBP-$N_{\Delta 21}^{344\text{-}SPSGS\text{-}348}$ were purified on amylose resin (New England Biolabs® Inc.), while P_{NTD} was purified as previously described (Leyrat et al., 2010). The $N_{\Delta 21}^{0}$-P_{NTD} complex was produced by incubating overnight at 4 °C an excess of P_{NTD} with MBP-$N_{\Delta 21}$-RNA complexes. This resulted in the appearance of MBP-$N_{\Delta 21}^{0}$-P_{NTD} complex that could be purified by size exclusion chromatography (SEC) on a S200 Superdex column (GE Healthcare®), checked by SDS-PAGE using a Precast 4-20 % gel (Biorad) (Figure 2B). The MBP moiety was cleaved using TEV protease overnight at 4°C. The resulting $N_{\Delta 21}^{0}$-P_{NTD} complex was then purified using Ni^{2+} chelate affinity chromatography followed by SEC. This procedure yielded pure $N_{\Delta 21}^{0}$-P_{NTD} complex.

Small Angle X-ray Scattering

The monodispersity of the samples used in SAXS experiments was checked by size exclusion chromatography (SEC) combined with detection by multi-angle laser light scattering (MALLS) and refractometry (RI) (Gérard, Ribeiro et al. 2007; Hodak, Wohlkönig et al. 2008). SAXS data were collected at the European Synchrotron Radiation Facility (Grenoble, France) on beamline ID14-3. The sample-to-detector distance was 1 m and the wavelength of the X-rays was 0.0995 nm. Samples were contained in a 1.9 mm wide quartz capillary. The time of exposure was optimized for reducing radiation damage. Data acquisition was performed at 20 °C. Protein concentrations ranged from 5 to 18 $mg.ml^{-1}$. Data reduction was performed using the established procedure available at ID14-3, and buffer background runs were subtracted from sample runs.

The radius of gyration and forward intensity at zero angle (*I(0)*) were determined with the programs PRIMUS (Konarev, Volkov et al. 2003) by using the Guinier approximation at low scattering vector values, Q ($Q = \frac{4\pi \sin\theta}{\lambda}$), in a $Q\,R_g$ range up to 1.3:

$$\ln I(Q) = \ln I(0) - \frac{R_g^2 Q^2}{3}$$

The forward scattering intensity was calibrated using bovine serum albumin and lysozyme as references. The radius of gyration and pairwise distance distribution function, *P(r)*, were calculated with the program GNOM (Semenyuk and Svergun 1991). The maximum dimension (D_{max}) value was adjusted so that the R_g value obtained from GNOM agreed with that obtained from the Guinier analysis.

Ab initio low-resolution shape reconstructions of the $N_{\Delta 21}$°-P_{NTD} complex was preformed from the scattering curves using the program DAMMIF (Franke and Svergun 2009). This program restore low-resolution shapes of the protein as a volume filled with densely packed spheres (dummy atoms) that reproduces the experimental scattering curve by a simulated annealing minimization procedure. DAMMIF minimizes the interfacial area between the molecule and the solvent by imposing compactness and connectivity constraints. 20 independent models were generated with DAMMIF with no symmetry restriction and were combined with the program DAMAVER (Volkov and Svergun 2003) yielding an average model that exhibits the common structural features of all the reconstructions.

X-ray Crystallography

Crystallization conditions were established using the Crystal Screen light (Hampton). After 24 hours of hanging drop vapor diffusion experiments, multiple needle-like crystals were observed in a droplet stored at 20°C containing 0.1µl of mother liquor 0.1M Sodium Acetate trihydrate and 4%w/v of PEG4000 pH 4.6 and 0.1µl of protein at a concentration of 8 mg/ml in Tris-HCl 20mM and 150mM NaCl buffer pH 7.5. By refining the crystallization conditions using slight variations in the mother liquor composition and droplet shape, we obtained crystals that grew to about 50 µm x 50 µm x 50 µm. A single $N_{\Delta 21}$°-P_{NTD} crystal was flash frozen at 100K in 25% glycerol and a 3.0 Å resolution diffraction data set was recorded at the wavelength 0.9334 Å at ID14 beamline in E.S.R.F., Grenoble. Diffraction data were processed using the program iMosflm (Powell 1999) and scaled with the program Scala from the ccp4 suite (Potterton, Briggs et al. 2003). The structure was solved by molecular replacement methods using the residues 22 to 422 of a protomer (2gic.pdb chain E) from the N-RNA crystal structure (Green, Zhang et al. 2006) as a search model in the program Phaser (McCoy, Grosse-Kunstleve et al. 2007). The program Buccaneer (Cowtan 2006) was used to trace and assign the sequence of P_{NTD}, and refinement of the structure was subsequently performed using Coot (Emsley, Lohkamp et al. 2010) and Refmac5 (Murshudov, Vagin et al. 1997). PROCHECK was used to check the quality of the model

(Laskowski, Macarthur et al. 1993). A table summarizing the crystallographic statistics is presented below (Table 1).

Space group	$P2_12_12$
Unit cell constants	a=74.56 Å b=171.97 Å c=239.86 Å α=β=γ=90°
Resolution limits[a]	59.97-3.03 Å (3.20-3.03 Å)
Number of measured reflections	168372
Number of unique reflections	54333
Completeness of data[a]	89.9% (80.7%)
Rmerge[a]	11.4% (28.1%)
Multiplicity[a]	3.1 (2.5)
I/σ[a]	9 (3.5)
R_{work}	24.2%
R_{free}	29.0%
Number of residues in favoured region Number of residues in allowed region Number of residues in outlier region	1974 (92.1%) 128 (6.0%) 37 (1.9%)
Rms BondLength	0.008
Rms BondAngle	0.966
Rms ChirVolume	0.067

[a] The values for the highest resolution shell are given in parentheses.

Structural modeling

Molecular dynamics simulations (MDS) with explicit solvent were performed using the GROMACS 4.0.2 software package (Hess, Kutzner et al. 2008) and the GROMOS 53a6 force field (Oostenbrink, Soares et al. 2005). At the beginning of each simulation, the protein was immersed in a box of SPC/E water molecules. Sodium and chloride ions were added to reach a 50mM salt concentration. Long range electrostatics was calculated with particle-mesh Ewald summation. Hydrogen atoms were treated as virtual sites, enabling a 5 fs integration time step to be used. The v-rescale thermostat (Bussi, Donadio et al. 2007) and the Parrinello-Rahman barostat (Rühle 2008) were used to maintain a temperature of 300 K and a pressure of 1atm. The system was equilibrated for 125 ps with restrained protein atoms before the beginning of the production simulation.

Protein docking was performed by using the program BiGGER (Krippahl, Moura et al. 2003). The receptor was composed of $N_{\Delta 21}$ bound to residues 17 to 31 of P_{NTD}, which constitute the α-helix present in the crystal structure and the ligand was made of residues 1 to 13 of P that were assumed to be α-helical based on P_{NTD} intrinsic conformational propensity (Leyrat et al., 2010). Three different representations of residues 1-13 of P_{NTD} that varied mainly at the side chains level were used to increase side chain conformational diversity. 30 different models of $N_{\Delta 21}$ bound to residues 17 to 31 of P_{NTD} were used that differed in the conformation of residues 343 to 363 of the C-loop of N. The multiple C-loop conformations were generated using the LOBO webserver (Tosatto, Bindewald et al. 2002). Cross-docking of these models with P_{NTD} residues 1-13 was conducted using BiGGER, generating 3 x 30 x 5000 potential solutions. The resulting decoys were analyzed in BiGGER and the 8 best ranked solutions according to BiGGER global score that displayed a sufficiently small distance (less than 1.6 nm) between residues 13 and 17 to allow connection with residues 14 to 16 were selected for further refinement using MDS. Residues 14 to 16 and 32 to 40 were modeled in using Coot. These models were simulated using the parameters described above and various parameters such as r.m.s.d., distances between the centre of mass of different groups of atoms, the number of atomic contacts and the percentage of helicity was monitored by using GROMACS routines (Figure 13). The models of $N_{\Delta 21}{}^0$-P_{NTD} obtained after 50 ns of MDS were compared to the SAXS data using the program CRYSOL (Svergun, Barberato et al. 1995). The combination of these parameters was used as a score to select the best model of the $N_{\Delta 21}{}^0$-P_{NTD} complex.

References

Albertini, A. A., A. K. Wernimont, et al. (2006). "Crystal structure of the rabies virus nucleoprotein-RNA complex." Science **313**(5785): 360-3.

Arnheiter, H., N. L. Davis, et al. (1985). "Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis." Cell **41**(1): 259-67.

Bussi, G., D. Donadio, et al. (2007). "Canonical sampling through velocity rescaling " J. Chem. Phys. **126**: 014101.

Chen, M., T. Ogino, et al. (2007). "Interaction of vesicular stomatitis virus P and N proteins: Identification of two overlapping domains at the N-terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA." J Virol **81**: 13478-13485.

Cowtan, K. D. (2006). "The Buccaneer software for automated model building." Acta Crystallographica Section D-Biological Crystallography **62**(9): 1002-1011.

Davis, N. L., H. Arnheiter, et al. (1986). "Vesicular stomatitis virus N and NS proteins form multiple complexes." J Virol **59**(3): 751-4.

De, B. P. and A. K. Banerjee (1984). "Specific interactions of vesicular stomatitis virus L and NS proteins with heterologous genome ribonucleoprotein template lead to mRNA synthesis in vitro." J Virol **51**(3): 628-34.

Emerson, S. U. and M. Schubert (1987). "Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus." Proc Natl Acad Sci U S A **84**(16): 5655-9.

Emerson, S. U. and Y. Yu (1975). "Both NS and L proteins are required for in vitro RNA synthesis by vesicular stomatitis virus." J Virol **15**(6): 1348-56.

Emsley, P., B. Lohkamp, et al. (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.

Fields, B. N., D. M. Knipe, et al. (1996). Fields' virology, 3rd edn. New York, Lippincott-Raven Publishers.

Franke, D. and D. I. Svergun (2009). "DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering." J. Appl. Cryst. **42**: 342-346.

Gérard, F., E. Ribeiro, et al. (2007). "Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution." Biochemistry **46**: 10328-10338.

Green, T. J., X. Zhang, et al. (2006). "Structure of the vesicular stomatitis virus nucleoprotein-RNA complex." Science **313**(5785): 357-60.

Hess, B., C. Kutzner, et al. (2008). "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable

Molecular Simulation." J. Chem. Theory Comput. **4**: 435-447.

Hill, V. M., L. Marnell, et al. (1981). "In vitro replication and assembly of vesicular stomatitis virus nucleocapsids." Virology **113**(1): 109-18.

Hodak, H., A. Wohlkönig, et al. (2008). "The peptidyl-prolyl isomerase and chaperone Par27 of *Bordetella pertussis* as the prototype for a new group of parvulins." J Mol Biol **376**(2): 414-26.

Howard, M. and G. Wertz (1989). "Vesicular stomatitis virus RNA replication: a role for the NS protein." J Gen Virol **70 (Pt 10)**: 2683-94.

Hudson, L. D., C. Condra, et al. (1986). "Cloning and expression of a viral phosphoprotein: structure suggests vesicular stomatitis virus NS may function by mimicking an RNA template." J Gen Virol **67 (Pt 8)**: 1571-9.

Konarev, P. V., V. V. Volkov, et al. (2003). "PRIMUS: a Windows PC-based system for small-angle scattering data analysis." J. Appl. Cryst. **36**: 1277-1282.

Krippahl, L., J. J. Moura, et al. (2003). "Modeling protein complexes with BiGGER." Proteins **52**(1): 19-23.

Laskowski, R. A., M. W. Macarthur, et al. (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." J. Appl. Cryst. **26**: 283-291.

Luo, M., T. J. Green, et al. (2007). "Conserved characteristics of the rhabdovirus nucleoprotein." Virus Res **129**(2): 246-51.

Masters, P. S. and A. K. Banerjee (1988). "Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA." J Virol **62**(8): 2658-64.

Mavrakis, M., F. Iseni, et al. (2003). "Isolation and characterisation of the rabies virus N degrees-P complex produced in insect cells." Virology **305**(2): 406-14.

Mavrakis, M., S. Mehouas, et al. (2006). "Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA." Virology **349**(2): 422-9.

McCoy, A. J., R. W. Grosse-Kunstleve, et al. (2007). "Phaser crystallographic software." J Appl Crystallogr **40**(Pt 4): 658-674.

Mellon, M. G. and S. U. Emerson (1978). "Rebinding of transcriptase components (L and NS proteins) to the nucleocapsid template of vesicular stomatitis virus." J Virol **27**(3): 560-7.

Murshudov, G. N., A. A. Vagin, et al. (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallographica Section D: Biological Crystallography **53**(3): 240-255.

Oostenbrink, C., T. A. Soares, et al. (2005). "Validation of the 53A6 GROMOS force field." Eur Biophys J **34**(4): 273-84.

Patton, J. T., N. L. Davis, et al. (1984). "N protein alone satisfies the requirement for protein synthesis during RNA replication of vesicular stomatitis virus." J Virol **49**(2): 303-9.

Peluso, R. W. (1988). "Kinetic, quantitative, and functional analysis of multiple forms of the vesicular stomatitis virus nucleocapsid protein in infected cells." J Virol **62**(8): 2799-807.

Peluso, R. W. and S. A. Moyer (1983). "Initiation and replication of vesicular stomatitis virus genome RNA in a cell-free system." Proc Natl Acad Sci U S A **80**(11): 3198-202.

Peluso, R. W. and S. A. Moyer (1988). "Viral proteins required for the in vitro replication of vesicular stomatitis virus defective interfering particle genome RNA." Virology **162**(2): 369-76.

Potterton, E., P. Briggs, et al. (2003). "A graphical user interface to the CCP4 program suite." Acta Crystallogr D Biol Crystallogr **59**(Pt 7): 1131-7.

Powell, H. R. (1999). "The Rossmann Fourier autoindexing algorithm in MOSFLM." Acta Crystallogr D Biol Crystallogr **55**(Pt 10): 1690-5.

Rühle, V. (2008). "Pressure coupling/barostats."

Semenyuk, A. V. and D. Svergun (1991). "GNOM - a program package for small-angle scattering data processing." J. Appl. Crystallog. **24**: 537-540.

Svergun, D., C. Barberato, et al. (1995). "CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates." J. Appl. Cryst. **28**: 768-773.

Tosatto, S. C., E. Bindewald, et al. (2002). "A divide and conquer approach to fast loop modeling." Protein Eng **15**(4): 279-86.

Volkov, V. V. and D. I. Svergun (2003). "Uniqueness of ab initio shape determination in small-angle scattering." J. Appl. Cryst. **36**: 860-864.

IV. Données supplémentaires:

A. Observation des protéines VSV N-MBP, $N_{\Delta 21}$-MBP et $N_{\Delta 21}{}^{344\text{-}SPSGS\text{-}348}$-MBP en microscopie électronique

Les complexes MBP-N-ARN, MBP-$N_{\Delta 21}$-ARN, et MBP-$N_{\Delta 21}{}^{344\text{-}SPSGS\text{-}348}$ du VSV produits dans *E.coli* et purifiés sur résine amylose (OZYME®) ont été observés en microscopie électronique afin de déterminer l'effet des mutations sur la morphologie des oligomères formés par la nucléoprotéine (Figure 33, Figure 34, et Figure 35, respectivement), et dont la présence a été mise en évidence par SEC-MALLS-RI (cf. ARTICLE VII) et par diffusion de lumière dynamique (données non présentées). Les clichés réalisés pour MBP-N-ARN révèlent la présence d'anneaux circulaires coexistant avec des formes linéaires plus longues de type nucléocapside et avec des aggrégats de taille variable (Figure 33).

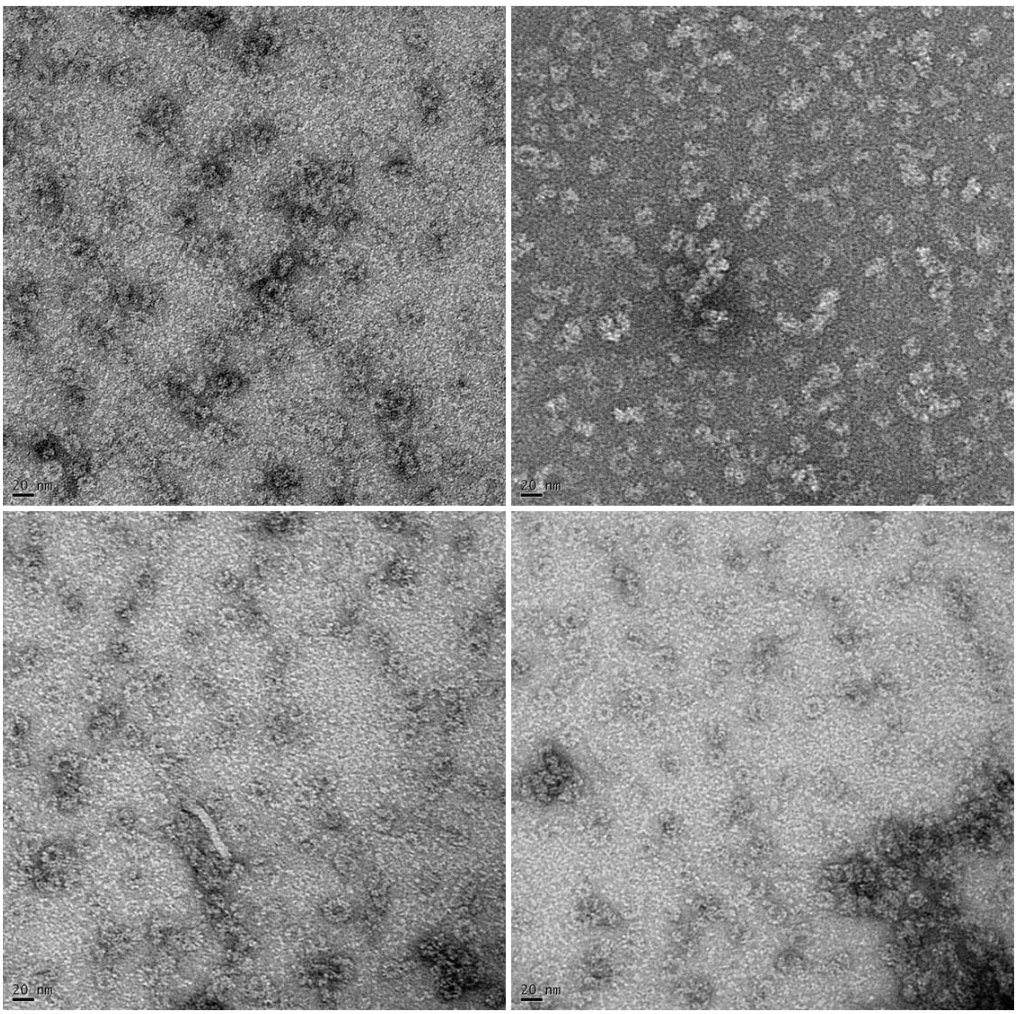


Figure 33 : clichés de microscopie électronique en coloration négative des complexes Nucléoprotéine-ARN-Maltose Binding Protein (N-ARN-MBP) du Virus de la Stomatite Vésiculaire. Les complexes ont été produits dans E. Coli et purifiés par affinité sur résine amylose. La barre d'échelle représente 20 nm. Les clichés ont été réalisés par Ambroise Desfosses.

Les images du complexe MBP-$N_{\Delta 21}$-ARN montrent une morphologie relativement similaire à ce qui est observé pour la forme complète de la nucléoprotéine, à l'exception d'une proportion plus importante d'aggrégats non circulaires de faible dimensions (Figure 34), suggérant une possible déstabilisation des structures oligomériques. Ce comportement est attendu du fait de la contribution significative des résidus 1 à 21 de la nucléoprotéine aux interactions N-N observées dans la structure cristallographique (Green, Zhang et al. 2006; Luo, Green et al. 2007).

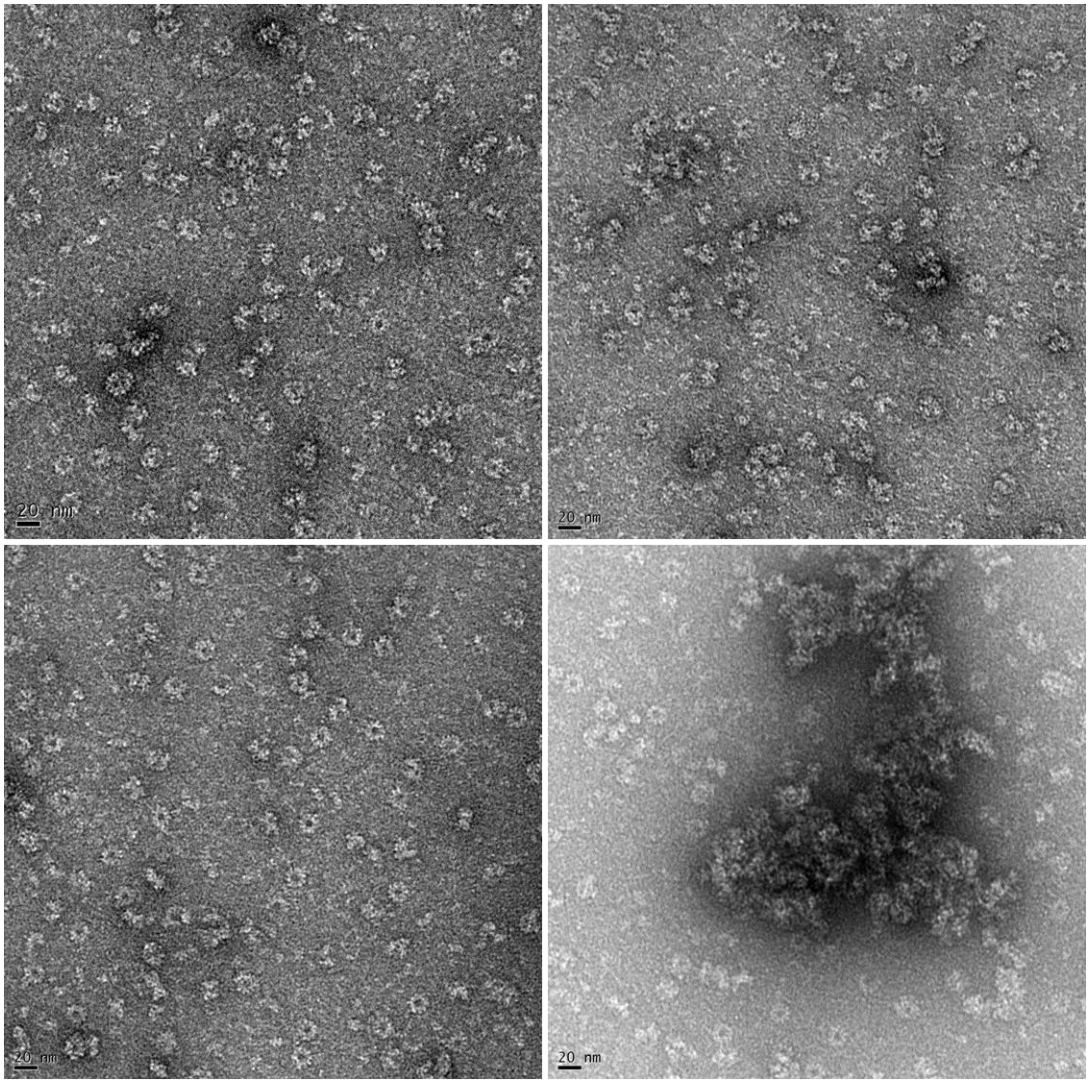


Figure 34 : clichés de microscopie électronique en coloration négative des complexes NucléoprotéineΔ21-ARN-Maltose Binding Protein ($N_{\Delta 21}$-ARN-MBP) du Virus de la Stomatite Vésiculaire. Les complexes ont été produits dans E. Coli et purifiés par affinité sur résine amylose. La barre d'échelle représente 20 nm. Les clichés ont été réalisés par Ambroise Desfosses.

Au contraire, les complexes MBP-$N_{\Delta 21}^{344\text{-}SPSGS\text{-}348}$, mutés au niveau des deux sous-domaines d'échange entre protomères en N- et en C-terminal semblent avoir perdus à la fois la capacité à former des oligomères circulaires (Figure 35), mais aussi toute capacité de fixation de

l'ARN. Ils existent néanmoins à l'état d'aggrégats solubles. Ces résultats indiquent probablement que l'oligomérisation et l'encapsidation d'ARN sont des processus couplés, comme cela a déjà été suggéré auparavant (Zhang, Green et al. 2008).

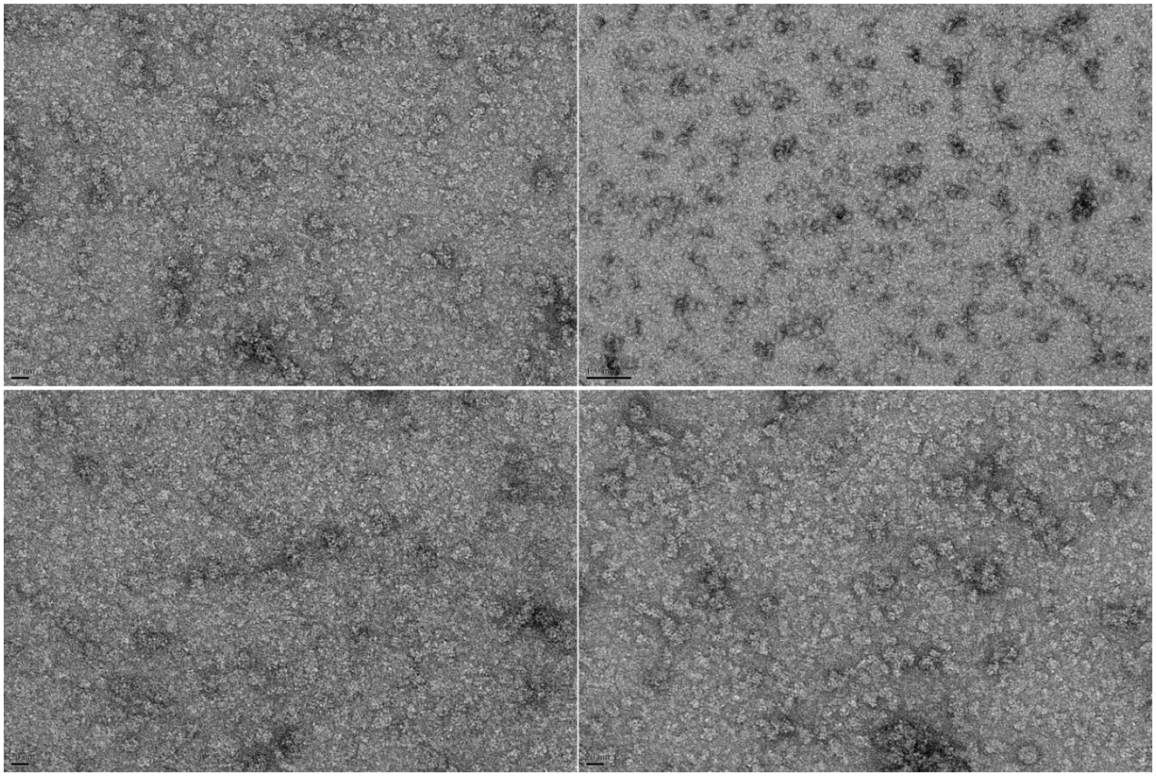


Figure 35 : clichés de microscopie électronique en coloration négative des complexes NucléoprotéineΔ21$^{344\text{-}SPSGS\text{-}348}$--Maltose Binding Protein ($N_{\Delta21}^{344\text{-}SPSGS\text{-}348}$- MBP) du Virus de la Stomatite Vésiculaire. Les complexes ont été produits dans E. Coli et purifiés par affinité sur résine amylose. La barre d'échelle est indiquée sur chaque image. Les clichés ont été réalisés par la plate-forme de microscopie électronique de l'IBS.

B. Mise en évidence de l'interaction entre les formes mutantes de la protéine N (MBP-$N_{\Delta21}$ et MBP-$N_{\Delta21}^{344\text{-}SPSGS\text{-}348}$) et la région P_{NTD} du VSV par RMN.

Les complexes MBP-$N_{\Delta21}$-ARN et MBP-$N_{\Delta21}^{344\text{-}SPSGS\text{-}348}$ ont été incubés pendant une heure avec un excès de P_{NTD} marqué ^{15}N et leurs spectres HSQC 2D ^{15}N- ^{1}H ont été comparés à une spectre correspondant au P_{NTD} isolé et enregistré dans des conditions identiques. Dans ces conditions (50mM Glu, 50mM Arg, 20 mM Tris-HCl, pH 6, T=10°C), les signaux correspondants à chaque résidu de P_{NTD} ont été atttribués (cf. Chap. III, ARTICLE II). Pour MBP-$N_{\Delta21}$-ARN comme pour MBP-$N_{\Delta21}^{344\text{-}SPSGS\text{-}348}$, on observe une diminution de tous les signaux lorsque la concentration de nucléoprotéine utilisée augmente, diminution qui peut être quantifiée et normalisée sous la forme d'un ratio I/I_0 (Figure 36). Il apparaît que la diminution de l'intensité des spots observés sur le spectre HSQC est plus marquée pour les résidus 1 à 15 de P_{NTD}, ce qui suggère un échange conformationnel sur une échelle de temps intermédiaire

(de l'ordre des micro- ou millisecondes). Cette observation est cohérente avec la diminution relative d'affinité pour P_{NTD} observée avec le mutant MBP-$N_{\Delta 21}^{344\text{-}SPSGS\text{-}348}$ et le modèle obtenu par docking flexible présenté dans la troisième partie du chapitre. En effet, la fixation des résidus 1 à 15 de P_{NTD} sur la boucle C-terminale désordonnée de N, qui n'établit pas d'interactions spécifiques avec le coeur du domaine N_{CTD}, pourrait être à l'origine de cette flexibilité accrue. Par ailleurs, l'impossibilité de former le complexe N^0-P à partir de la forme MBP-N-ARN indique que l'accessibilité de cette boucle C-terminale au solvant est requise pour sa formation. Cette expérience montre également que la forme mutante MBP-$N_{\Delta 21}^{344\text{-}SPSGS\text{-}348}$ est capable d'interagir efficacement avec P_{NTD}, et confirme donc que la diminution d'affinité relative observée par SEC-MALLS-RI (cf. partie III, ARTICLE VII) est effectivement liée à une interaction moins efficace entre la boucle C-terminale de N et P_{NTD} et non à un défaut de repliement protéique causé par les mutations introduites.

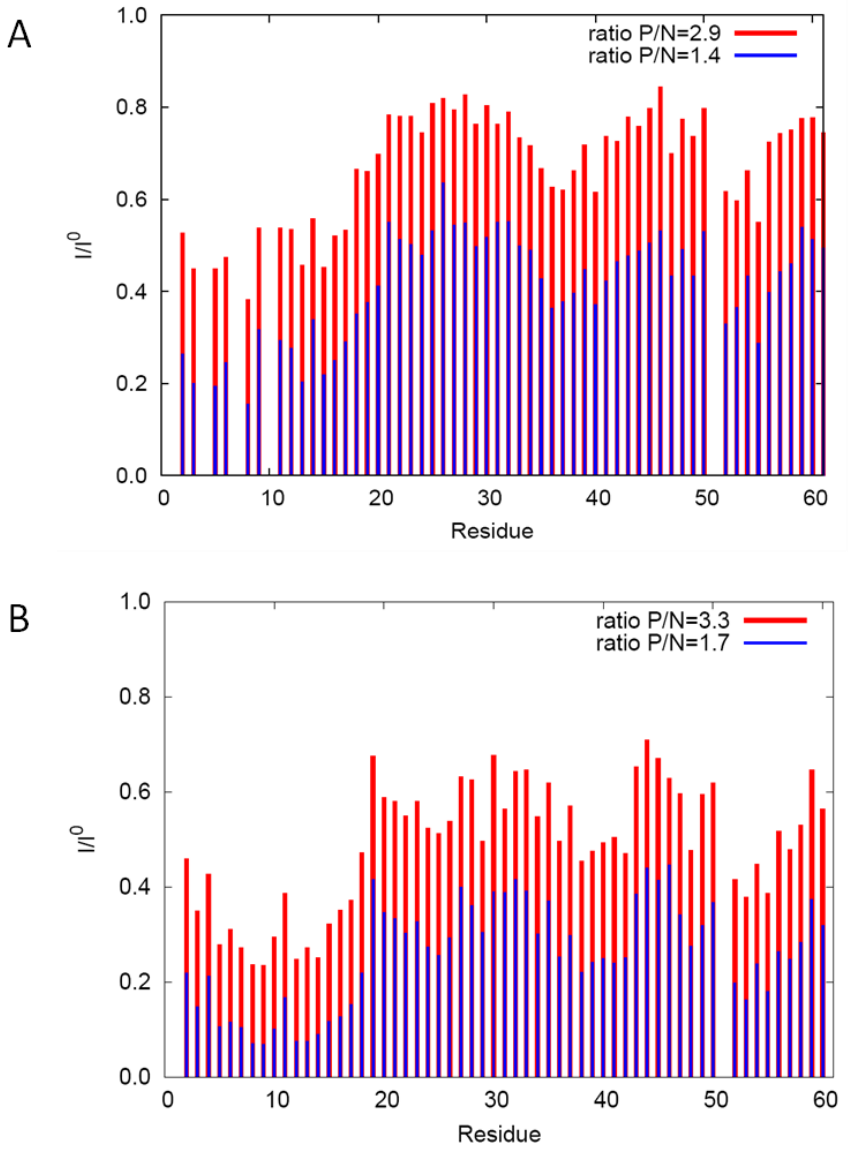


Figure 36 : Titration de VSV P_{NTD} par MBP-$N_{\Delta 21}$ (A) et MBP-$N_{\Delta 21}^{344\text{-}SPSGS\text{-}348}$ (B) par RMN. La diminution de l'intensité des signaux des Cα précédemment attribués du spectre HSQC 2D ^{15}N- ^{1}H de P_{NTD} entraînée par l'augmentation de la concentration de $N_{\Delta 21}$- MBP est représentée sous la forme d'un rapport $I/I_{0,}$ I_0 correspondant à l'intensité mesurée en absence de MBP-$N_{\Delta 21}$.

C. Production et caractérisation structurale du complexe RV $N_{\Delta 23}$-P_{1-68}.

Le complexe MBP-$N_{\Delta 23}$-ARN du RV a été produit en utilisant une procédure identique à celle décrite pour VSV (cf. partie III, ARTICLE VII). L'observation du complexe en microscopie électronique suggère une morphologie similaire au complexe MBP-$N_{\Delta 21}$-ARN du VSV, à l'exception d'une proportion relative d'anneaux moins importante dans la préparation (Figure 37).

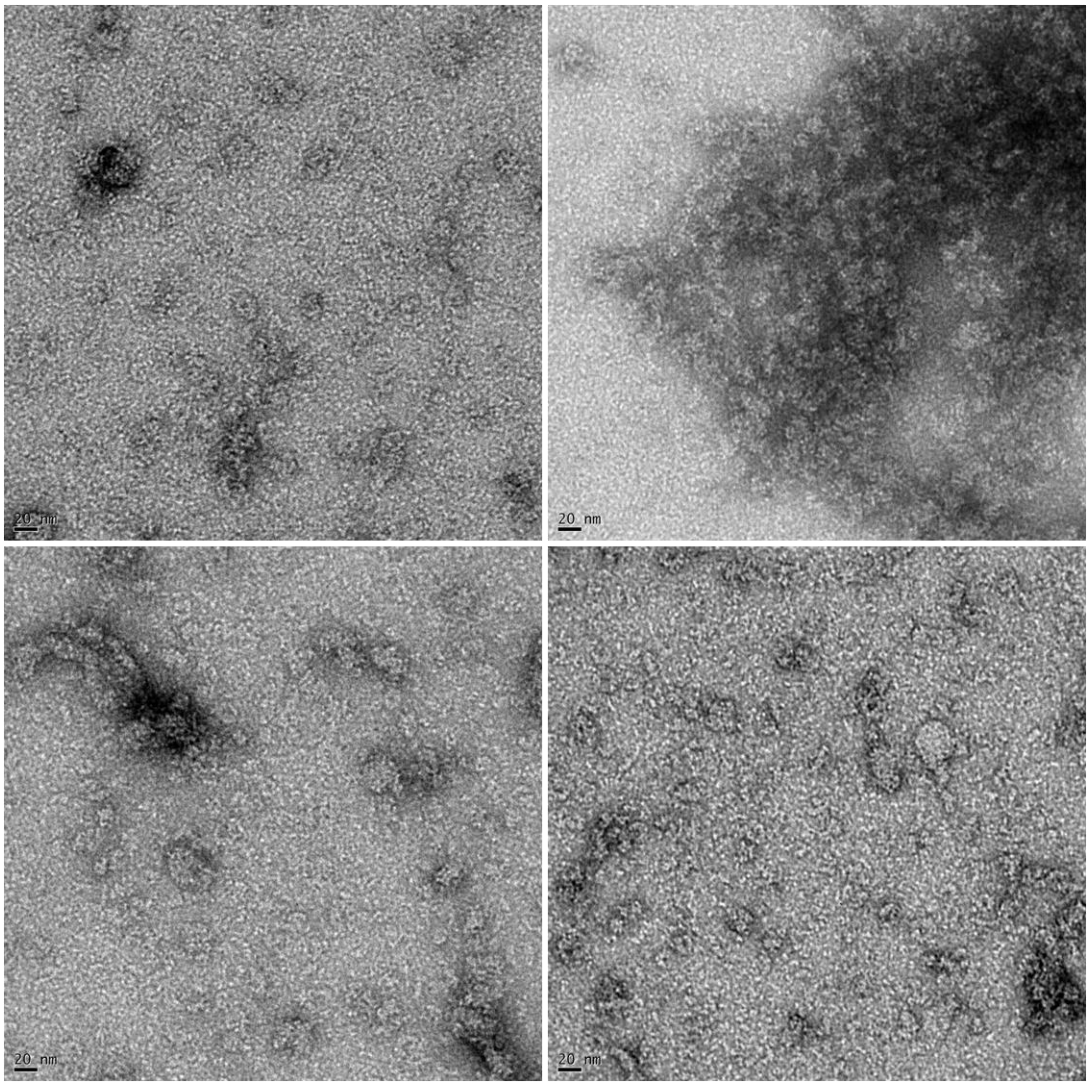


Figure 37 : clichés de microscopie électronique en coloration négative des complexes Maltose Binding Protein-NucléoprotéineΔ23-ARN (MBP-$N_{\Delta 23}$-ARN) du RV. Les complexes ont été produits dans E. Coli et purifiés par affinité sur résine amylose. La barre d'échelle représente 20 nm. Les clichés ont été réalisés par Ambroise Desfosses.

La présence d'oligomères de taille importante est aussi apparente sur le profil enregistré en SEC-MALLS-RI, qui montre notamment que les complexes MBP-$N_{\Delta 23}$-ARN sont élués à un volume inférieur à 13 ml, et en partie de le volume mort de la colonne Superdex 200 (GE Healthcare®) (Figure 38A), indiquant un rayon hydrodynamique moyen

(R_S) supérieur à 6 nm. En effet, une analyse par diffusion de lumière dynamique (DLS) indique un R_S d'environ 15 nm. Par ailleurs, un pic minoritaire correspondant à la MBP libre autoclivée (possiblement mélangé à une faible quantité de protéine $N_{\Delta 23}^{0}$ autoclivée) est visible à un volume $V_{élution}$ ~ 16 ml. Tout comme pour le VSV, l'incubation des complexes MBP-$N_{\Delta 23}$-ARN avec P_{NTD} (résidus 1-68) entraîne l'apparition d'un pic de complexe MBP-$N_{\Delta 23}$-P_{NTD} (Figure 38B), dont la masse moléculaire mesurée par MALLS après purification est d'environ 100 kDa (Figure 38C), en accord avec l'association d'un monomère de nucléoprotéine fusionné à la MBP (90 kDa) avec un P_{NTD} (9 kDa). La stoechiométrie 1 :1 du complexe est confirmée par l'analyse SEC-MALLS-RI du complexe clivé et purifié $N_{\Delta 23}$-P_{NTD}, qui indique une masse moléculaire d'environ 55kDa (Figure 38D), en accord avec la masse calculée d'après la séquence (58kDa).

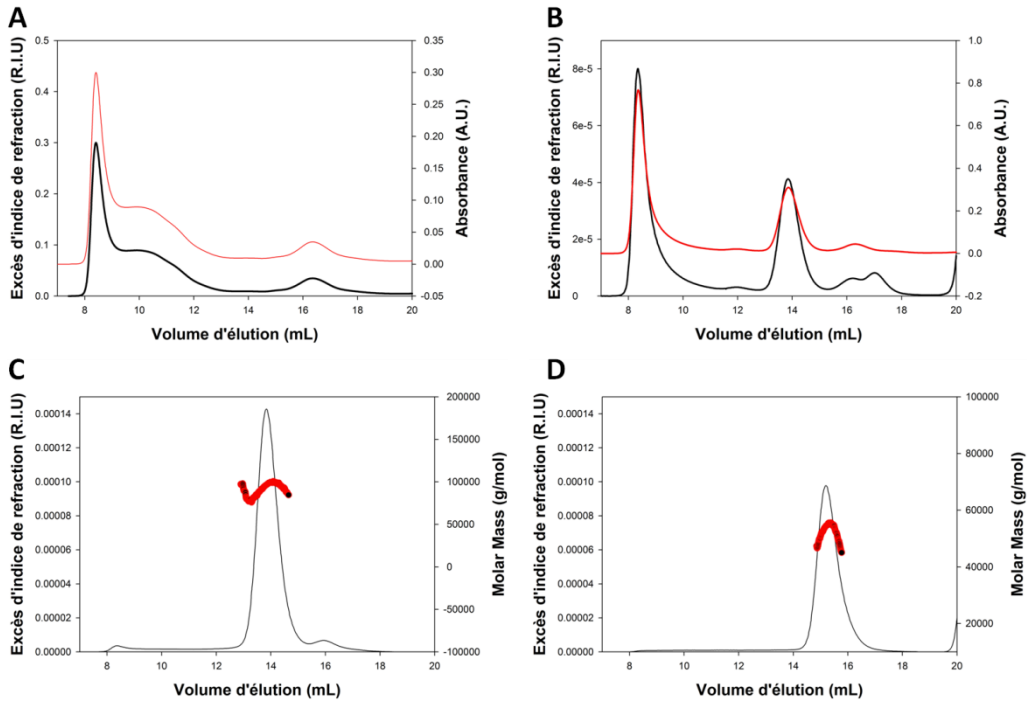


Figure 38 : Formation du complexe $N_{\Delta 23}^{0}$-P_{NTD} chez le Virus de la Rage. A et B. Profiles SEC-RI obtenus pour les complexes NucléoprotéineΔ23-ARN-Maltose Binding Protein ($N_{\Delta 23}$-ARN-MBP) avant et après 1 heure d'incubation avec P_{NTD} (résidus 1 à 68), respectivement. La courbe noire représente le signal de refractométrie et la courbe rouge représente l'absorbance à 280 nm. **C et D.** Profiles SEC-MALLS-RI obtenus pour les complexes $N_{\Delta 23}^{0}$-MBP-P_{NTD} et $N_{\Delta 23}^{0}$ -P_{NTD}, respectivement après purification sur une colonne Superdex 200 (GE Healthcare®). La courbe noire représente le signal de refractométrie et les cercles rouges représentent la masse moléculaire.

Le complexe $N_{\Delta 23}^{0}$-P_{NTD} a été analysé par SAXS sur la ligne de lumière ID14-3 (ESRF, Grenoble). Les profiles de diffusion enregistrés à trois concentrations (3, 6 et 10 mg/ml), et pour des valeurs de Q comprises entre 0.02 nm^{-1}< Q < 5.90 nm^{-1} sont représentés sur la Figure 39A. L'analyse de Guinier indique un rayon de giration (R_g) de 3.0 nm et l'absence d'agrégation ou d'effets intermoléculaires dans la gamme de concentration utilisée (Figure 39B). La forme du graphe de Kratky (Figure 39C), ainsi que la légère extension observée sur la fonction P(r) (Figure 39D) suggèrent que le complexe est globalement replié

mais possède cependant une forme légèrement allongée et/ou une faible proportion de régions désordonnées.

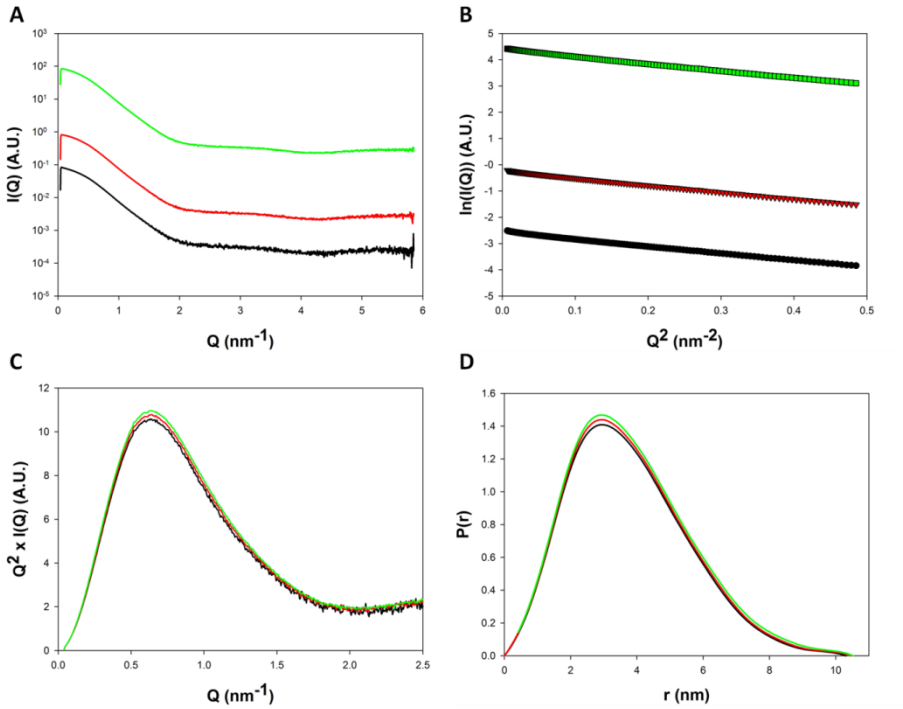


Figure 39 : Expériences de diffusion de rayons X aux petits angles (SAXS) sur le complexe $N_{\Delta 23}$-P_{NTD} du Virus de la Rage. A. Les données de SAXS ont été enregistrées pour des valeurs de Q comprises entre 0.02 nm^{-1}< Q < 5.90 nm^{-1}. Les courbes obtenues à trios concentrations différentes (3, 6 et 10 mg/ml) sont représentées en noir, en rouge, et en vert, respectivement. **B.** Analyse de Guinier réalisées dans la région $Q.R_g < 1.3$. **C.** Graphes de Kratky aux trois concentrations indiquant que le complexe est globalement replié. **D.** Fonctions de distribution de distances (pair distribution function, P(r)).

La fonction de distribution de distance des paires interatomiques P(r) a été utilisée pour reconstituer *ab initio* la forme basse résolution du complexe $N_{\Delta 23}{}^{0}$-P_{NTD} (Figure 40) à l'aide du programme DAMMIF (Franke and Svergun 2009). L'enveloppe obtenue est relativement similaire à ce qui a été observé chez le VSV, et rappelle la forme d'un protomère de nucléoprotéine ayant sa cavité de fixation à l'ARN remplie de densité électronique. Les différences avec VSV consistent principalement en un D_{max} légèrement plus élevé (10.5 nm contre 9.3 nm), qui pourrait résulter de la longueur plus importante de la boucle désordonnée C-terminale, et à la présence d'une densité légèrement plus volumineuse au niveau de la cavité de fixation à l'ARN, semblant se localiser sur une des faces latérales de la nucléoprotéine et qui pourrait correspondre aux 8 résidus supplémentaires inclus dans la construction du P_{NTD} chez RV. La Figure 40 montre une superposition de l'enveloppe basse résolution moyenne du complexe $N_{\Delta 23}{}^{0}$-P_{NTD} avec un des protomères cristallographiques (code pdb : 2GTT) dans lequel la boucle C-terminale a été modélisée (Ribeiro, Leyrat et al. 2009) et avec la structure de ce même protomère après 5 ns de simulation de dynamique moléculaire en solvant explicite. L'alignement structural souligne la flexibilité des longues boucles N- et C-terminales de la nucléoprotéine, et en particulier de la boucle C-terminale, qui occupe une conformation différente dans les trois modèles (modèle atomique à l'état

cristallin, modèle atomique théorique extrait de la simulation, et enveloppe basse résolution dérivée des données SAXS).

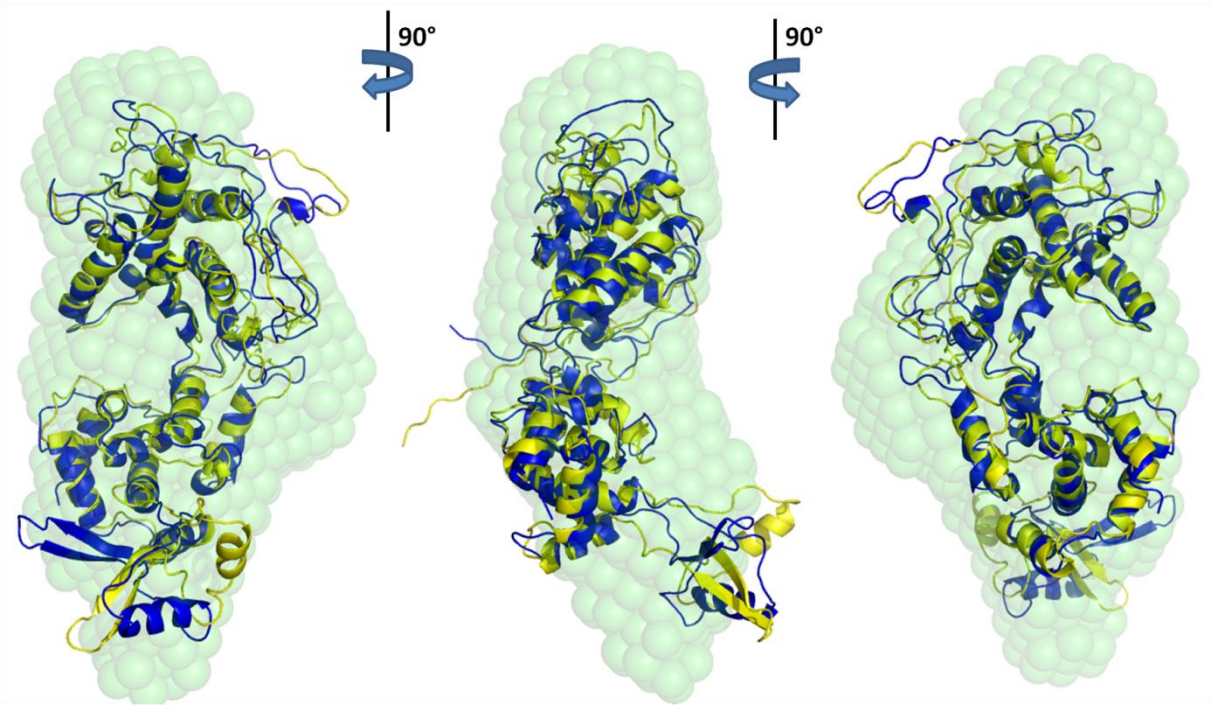


Figure 40 : Enveloppe basse résolution du complexe $N_{\Delta 23}$-P_{NTD} du Virus de la Rage. La structure basse résolution correspondant à la moyennation de 20 modèles indépendants généré par DAMMIF est superposée à deux structures de nucléoprotéine tronquées au niveau du résidu 22, et qui correspondent à un protomère extrait de la structure cristallographique du complexe N-ARN (code pdb : 2GTT) dans lequel la boucle C-terminale désordonnée a été modélisée (Ribeiro, Leyrat et al. 2009) et à un modèle extrait d'une simulation de dynamique moléculaire réalisée sur N^0 à t = 5 ns, respectivement en jaune et en bleu. Trois vues différant par une rotation de 90° sont représentées.

L'incubation d'un excès de complexe $N_{\Delta 23}{}^0$-P_{NTD} avec la phosphoprotéine (mutant ΔC261) du RV (Figure 41A) conduit à la formation de complexe $N_{\Delta 23}{}^0$-P entraînant un léger déplacement du pic d'élution de la phosphoprotéine (Figure 41A). L'analyse du contenu du pic par SDS-PAGE indique qu'il est constitué par l'association d'un dimère de phosphoprotéine (67 kDa) sur chaque monomère $N_{\Delta 23}$ (49 kDa) (Figure 41C), ce qui est confirmé par l'analyse SEC-MALLS-RI du complexe purifié fournissant une masse moléculaire estimée d'environ 110 kDa en accord avec la masse calculée d'après la séquence (116 kDa) (Figure 41B).

Des données de SAXS ont également été enregistrées à trois concentrations (0.1, 0.3 et 0.6 mg/ml) pour ce complexe (Figure 42A), et ce en utilisant directement les fractions récoltées à l'issue de la purification par SEC et analysées par SDS-PAGE et MALLS (Figure 41). L'analyse de Guinier indique un rayon de giration (R_g) de 5.4 nm et, une fois encore, l'absence d'agrégation ou d'effets intermoléculaires dans la gamme de concentration utilisée (Figure 42B). La forme du graphe de Kratky (Figure 42C), ainsi que l'extension marquée de la fonction P(r) (Figure 42D) sont typiques des protéines partiellement désordonnées et ressemblent à ce qui est observé pour les phosphoprotéines isolées de RV et VSV (Chap.V).

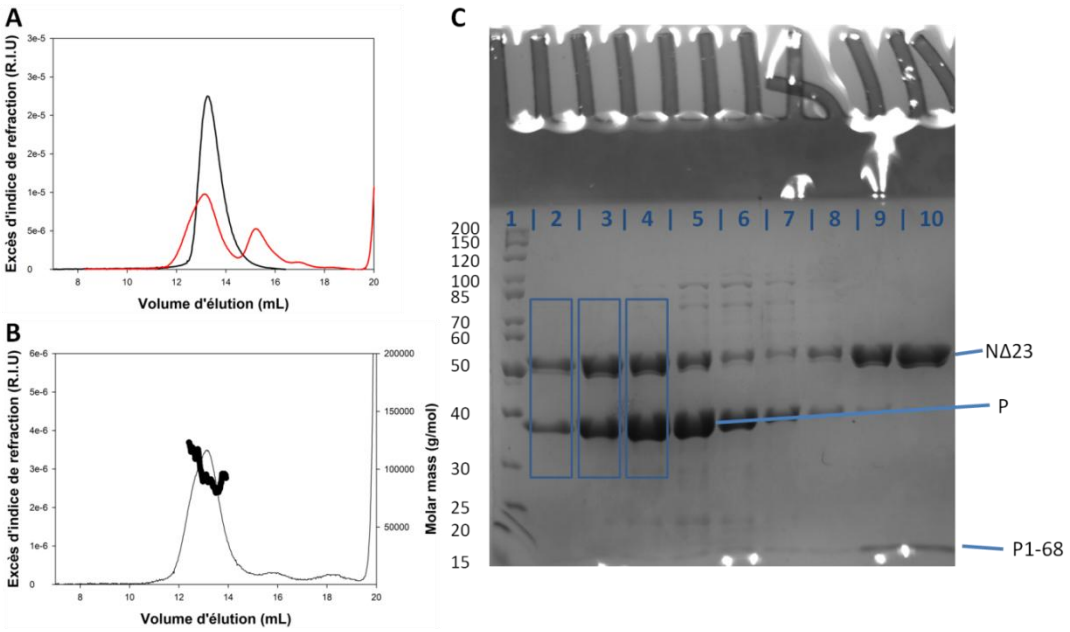


Figure 41 : Formation du complexe $N_{\Delta 23}^0$-P chez le Virus de la Rage. A. Profiles SEC- -RI obtenus pour la phosphoprotéine (courbe noire) et un mélange constitué de phosphoprotéine incubée une heure avec un excès de complexe $N_{\Delta 23}^0$-P_{NTD}. **B.** Profil SEC-MALLS-RI obtenu pour le complexe $N_{\Delta 23}^0$-P, après purification sur une colonne Superdex 200 (GE Healthcare®) et correspondant à la fraction 4 analysée par SDS-PAGE en C. **C.** Analyse SDS-PAGE (12 %) du profil d'élution représenté en A sur colonne S200 et montrant les complexes $N_{\Delta 23}^0$-P (fractions 2 à 7) et $N_{\Delta 23}^0$-P_{NTD} (fractions 7 à 10). Les fractions utilisées pour les expériences de SAXS (2, 3 et 4) sont encadrées en bleu.

Il est intéressant de noter que l'augmentation de R_g observée pour le complexe $N_{\Delta 23}^0$-P_2 par rapport à la protéine P seule est faible (5.4 contre 5.1 nm), tout comme l'augmentation de R_s à l'origine de la faible différence de volume d'élution observée entre $N_{\Delta 23}^0$-P_2 et P_2. Ceci est également valable en ce qui concerne les dimensions maximales de l'objet (D_{max}). Cette faible augmentation des dimensions de la molécule suggère que l'association de P_2 à N^0 limite la flexibilité de la phosphoprotéine grâce à l'interaction de P_{NTD} avec la cavité de fixation à l'ARN de N^0, mais aussi peut-être *via* une interaction de P_{CTD} avec N^0, l'étape de fixation monomoléculaire étant probablement favorisée une fois la région P_{NTD} déjà ancrée sur la nucléoprotéine. La modélisation *ab initio* du complexe à partir des données SAXS produit des enveloppes basse résolution aux formes allongées, qui sont similaires à celle obtenues à partir des données SAXS de la phosphoprotéine seule (Chap.V, Figure 26), mais possèdent cependant une région plus dense dont la forme rappelle celle d'une nucléoprotéine, et semblent être ancrés au reste de la structure par deux points d'ancrage (données non présentées). Cette particularité est également observée pour le complexe $N_{\Delta 21}^0$-P_2 du VSV, pour lequel des données SAXS ont également été enregistrées (données non présentées).

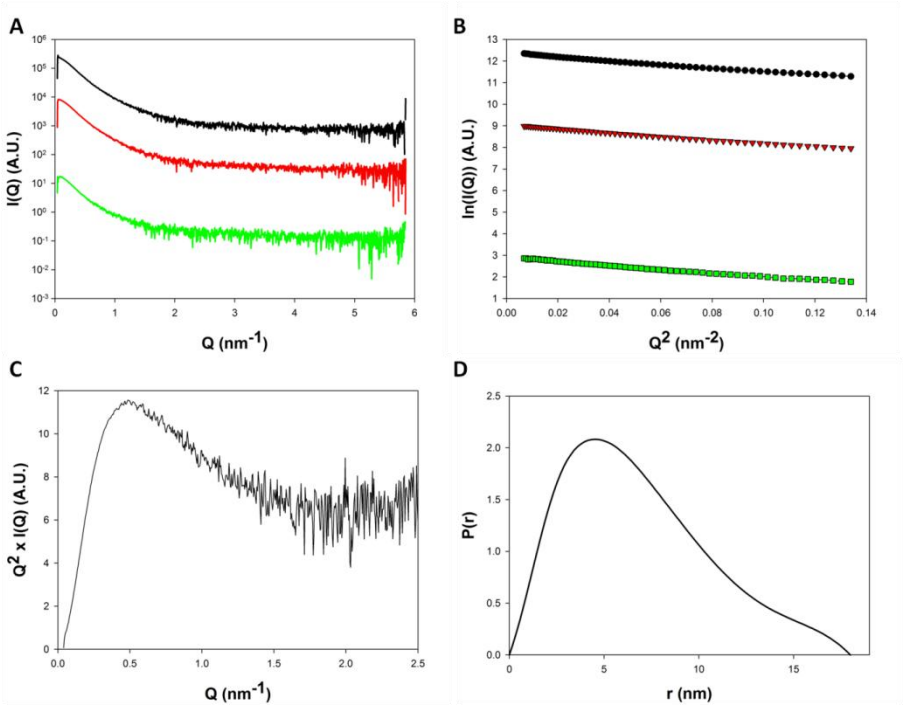


Figure 42 : Expériences de diffusion de rayons X aux petits angles (SAXS) sur le complexe $N_{\Delta 23}^{0}$-P du Virus de la Rage. A. Les données de SAXS ont été enregistrées pour des valeurs de Q comprises entre 0.05 nm^{-1}< Q < 5.90 nm^{-1}. Les courbes obtenues à trois concentrations différentes (0.1, 0.3, et 0.6 mg/ml) sont représentées en vert, en rouge, et en noir, respectivement. **B.** Analyse de Guinier réalisée dans la région $Q.R_g$ < 0.8. **C.** Graphe de Kratky indiquant que le complexe possède une composante désordonnée significative. **D.** Fonction de distribution de distances (pair distribution function, P(r)).

La modélisation atomique des données SAXS des complexes $N_{\Delta 21}^{0}$-P_2 de VSV et $N_{\Delta 23}^{0}$-P_2 de RV sous la forme d'un ensemble de structures, comme cela a été réalisé pour la phosphoprotéine P_2 du VSV (Chap. V), permettrait d'explorer l'espace conformationnel du complexe N^0-P_2 des *Rhabdoviridae* et autoriserait des comparaisons intéressantes entre P_2 et N^0-P_2 d'une part, et entre VSV et RV d'autre part.

D. Interprétation à posteriori des données de double hybride chez la levure concernant l'interaction entre la protéine N et la région P_{NTD} chez le VSV et chez RV

L'analyse par double hybride chez la levure *S. cerevisae* de l'interaction entre des mutants de la nucléoprotéine comportant diverses substitutions avec la phosphoprotéine ou un mutant déleté de son domaine C-terminal avait pour but initial de valider expérimentalement la prédiction bioinformatique du site de fixation de P_{NTD} sur N^0, c'est-à-dire la cavité de fixation à l'ARN. Pour ce faire, nous avons muté des résidus de N impliqués dans la fixation à l'ARN et fréquemment impliqués dans des interactions électrostatiques avec P_{NTD} dans les modèles du complexe obtenus par arrimage. Ce travail a été réalisé en collaboration avec le laboratoire du Dr. Danielle Blondel (LVMS, Gif-Sur-Yvette). La Figure 43 montre les résultats obtenus en remplaçant les résidus Arg143, Arg146 et Arg317 par des alanines dans une série de simples et de doubles mutants de N chez le VSV. Bien qu'aucune différence significative de signal ne soit observée pour les complexes formés entre N et P ou entre N et P1-184, et ce

pour la forme sauvage de N et les trois mutants R143A, R146A et R317A, on observe une perte complète de l'interaction chez deux des doubles mutants (R143A/R317A et R146A/R317A).

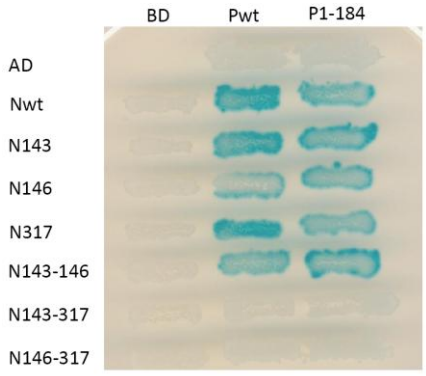


Figure 43: Interaction des protéines N et P du VSV par double hybride: L'interaction entre P ou la forme tronquée P1-184 fusionnés au domaine de fixation à l'ADN (BD) (colonnes) avec des mutants de N fusionnés à GAL4 AD (lignes) a été mesurée par l'apparition de colonies bleues en présence de X-Gal.

L'analyse de l'effet de ces mutations sur base de la structure cristallographique révèle à *posteriori* que les résidus Arg146 et Arg317 ne sont pas localisés directement à l'interface entre les deux protéines, et que seul Arg143 pourrait éventuellement constituer un « hot spot » ($\Delta\Delta G_{binding}$ = 1.5 kcal/mol prédit par FoldX (Schymkowitz, Borg et al. 2005)). En revanche, bien que la perte de l'interaction observée pour R143A/R317A et R146A/R317A puisse provenir d'un problème lié au repliement des protéines N doublement mutées dans la levure, il est également possible que la suppression simultanée de deux charges positives au sein de la cavité de fixation à l'ARN réduise l'affinité de N pour P et/ou entraîne des changements conformationnels de la nucléoprotéine aboutissant à la destabilisation du complexe N^0P.

Des expériences de double hybride similaires ont été réalisées chez le RV. Cependant, les mutants de N réalisés comportaient des inversions de charge (R168E, R323E, R434E, et R168E/R323E) supposées destabiliser d'éventuels ponts salins entre N^0 et P_{NTD} ou des substitutions en tryptophane censés introduire une gêne stérique s'opposant à la formation du complexe (V230W et S237W). Les résultats obtenus sont présentés dans la Figure 44. La mutation R168E semble être sans effet sur l'interaction, alors que les substitutions R323E, R434E et V230W tendent à augmenter l'interaction entre PΔC125 et N, suggérant un effet spécifique sur la formation du complexe N^0-P. Au contraire, la capacité du mutant S237W à s'associer à PΔC125 est totalement abolie par la mutation, alors que l'intensité du signal correspondant à la fixation aux complexes N-ARN demeure intacte, ce qui ici encore suggère un effet spécifique de la mutation sur la formation du complexe N^0-P. Le double mutant R168E/R323E perd toute interaction avec la phosphoprotéine. Comme dans le cas du VSV, la double mutation induit une modification importante de la charge nette de la cavité de fixation à l'ARN de la nucléoprotéine, qui est ici amplifiée par l'inversion de charge qui modifie la charge de 4 unités, entraînant potentiellement des changements conformationnels et une baisse d'affinité pour la région P_{NTD} fortement chargée négativement.

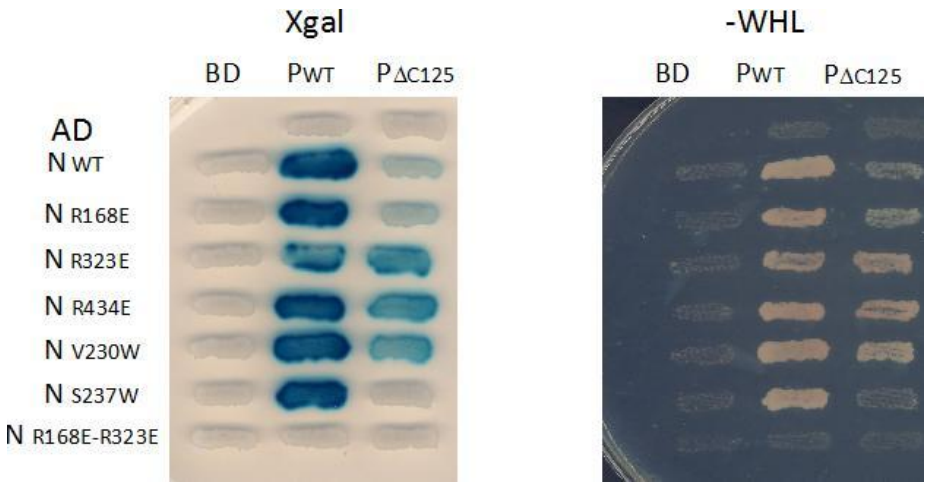


Figure 44: Interaction des protéines N et P de RV par double hybride: L'interaction entre P ou la forme tronquée P1-184 fusionnés au domaine de fixation à l'ADN (BD) (colonnes) avec des mutants de N fusionnés à GAL4 AD (lignes) a été mesurée par l'apparition de colonies bleues en présence de X-Gal (à gauche), ou par une croissance réduite des levures cultivées sur milieu SD déficient en Trp, His et Leu (à droite).

Un alignement structural a été réalisé entre la sous-unité monomérique de $N_{\Delta 21}{}^{0}$-P_{NTD} présente dans le cristal chez le VSV (cf. Partie III, ARTICLE VII) et un protomère de la nucléoprotéine du RV extrait de la structure cristallographique du complexe N-ARN (code pdb : 2GTT) (Figure 45). En admettant un site d'interaction similaire pour P_{NTD} dans les deux systèmes viraux, les données de double hybride concernant les mutants de nucléoprotéine peuvent être réinterprétées. Il apparaît d'après cette analyse que les résidus mutés sont localisés dans l'interface d'interaction, à l'exception du résidu Arg168, ce qui explique l'absence d'effet observé en mutant ce résidu. L'augmentation du signal mesuré en double hybride pour les mutants R323E et R434E suggère que ces résidus chargés interagissant avec l'ARN dans le complexe N-ARN ne sont pas impliqués dans des ponts salins avec P_{NTD}. La mutation V230W entraîne également une augmentation apparente de l'affinité pour P_{NTD}. Cependant, ce résidu hydrophobe est conservé entre VSV et RV et aligné structuralement avec le résidu Val219 de VSV N. Il est de plus localisé dans une région hydrophobe de l'interface et sa mutation en tryptophane est susceptible de renforcer la formation d'une surface de contact hydrophobe entre N et P_{NTD}. En revanche, la mutation S237W introduit clairement une gêne stérique au niveau de l'interface avec P_{NTD} (Figure 45), de manière consistante avec la perte spécifique de l'interaction avec PΔC125 observée en double hybride.

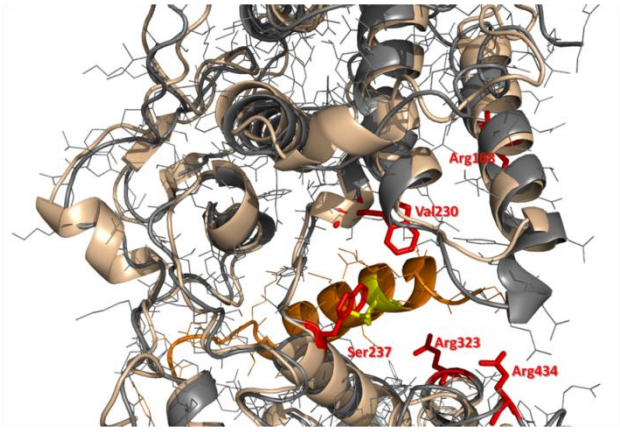


Figure 45 : Alignement structural entre le complexe $N_{\Delta 21}{}^{0}$-P_{NTD} cristallographique du VSV (Partie III, ARTICLE VII) et un protomère de nucléoprotéine extrait du complexe N-ARN du Virus de la Rage (code pdb : 2GTT). Les structures sont représentées en cartoons (couleur blé pour VSV N, gris 50 % pour RV N, et orange pour VSV P_{NTD}). Les résidus mutés dans la nucléoprotéine sont représentés en bâtonnets rouge, et le résidu de P_{NTD} responsable de la gêne stérique avec N S237W est coloré en jaune (VSV P_{NTD} Glu28).

V. Conclusion

L'étude structurale du complexe N^0-P par SEC-MALLS-RI, SAXS, RMN et cristallographie des rayons X (Parties III&IV), rendue expérimentalement possible par l'utilisation des mutants MBP-$N_{\Delta 21}$ chez le VSV et MBP-$N_{\Delta 23}$ chez RV, nous a permis de confirmer les observations précédemment consignées dans la littérature concernant la stœchiométrie du complexe, sa forme allongée, et l'implication des résidus de la région N-terminale de la phosphoprotéine dans l'interaction (Mavrakis, Iseni et al. 2003; Mavrakis, Mehouas et al. 2006; Chen, Ogino et al. 2007), mais également de valider l'implication de la cavité centrale de la nucléoprotéine dans l'interaction, inférée d'après les résultats de docking (Partie II). La structure cristallographique du complexe $N_{\Delta 21}{}^0$-P_{NTD} du VSV nous a fournit des informations importantes sur l'activité de chaperonne de la phosphoprotéine vis-à-vis de la protéine N. En particulier, le site de fixation de P_{NTD} sur N^0 recouvre partiellement le site de fixation de l'ARN et empêche l'association stable des protomères de nucléoprotéine en gênant stériquement le positionnement du sous-domaine N-terminal composé des résidus 1 à 21 de N.

Les données complémentaires concernant la caractérisation du mutant MBP-$N_{\Delta 21}{}^{344\text{-SPSGS-}348}$, l'impossibilité de former le complexe N^0-P à partir des complexes N-ARN natifs, l'existence du complexe $N_{\Delta 21}{}^0$-P_{NTD} à l'état monomérique en solution confirmée par SEC-MALLS-RI et par SAXS, les expériences de titration par RMN, mais aussi les modélisations *ab initio* du complexe, une simulation de dynamique moléculaire réalisée à partir de la sous-unité cristallographique, ainsi que la modélisation de l'interaction entre la boucle C-terminale de N et les résidus N-terminaux de P_{NTD} par docking flexible suggèrent que le complexe N^0-P est additionnellement stabilisé par l'association des résidus 1 à 12 de la phosphoprotéine avec la région de la boucle C-terminale de N impliquée dans la formation de contacts N-N (résidus 342-350).

L'étude SAXS du complexe $N_{\Delta 23}{}^0$-P_{NTD} chez le RV, ainsi que l'interprétation à *posteriori* de données d'interaction par double hybride suggèrent que la structure du complexe N^0-P est globalement similaire chez le VSV et chez RV. Cependant, des différences significatives semblent exister, et l'obtention d'une structure haute résolution permettrait une comparaison fine entre les deux systèmes. Malheureusement, plusieurs essais de cristallisation ont été réalisés sans succès jusqu'à présent (données non présentées), et la cristallisation du complexe nécessitera probablement l'utilisation de constructions plus courtes de P_{NTD}, voire l'introduction de mutations destinées à favoriser l'empilement cristallin (Derewenda 2004; Derewenda and Vekilov 2006).

CHAPITRE VIII: DISCUSSION & PERSPECTIVES

gcq#285: "Baby, It Aint Over Till It's Over"

(Lenny Kravitz)

Structure modulaire de la phosphoprotéine

Les prédictions bioinformatiques sur base des séquences en acides aminés, combinées à la caractérisation biophysique et structurale des domaines structurés ainsi identifiés ont montré que la phosphoprotéine des *Rhabdoviridae* possède une organisation structurale conservée caractérisée par la présence de deux domaines autonomes repliés (P_{CED} et P_{CTD}) séparés par deux longues régions désordonnées (IDR_{NTD} et IDR_{CTD}). Trois modules fonctionnels (P_{NTD}, P_{CED} et P_{CTD}) impliqués respectivement dans la fixation à N^0 et à L, l'homodimérisation de la protéine, et la fixation aux complexes N-ARN sont particulièrement conservés (cf. Chap.II, ARTICLE I). Ces trois modules fonctionnels sont représentés dans la Figure 46A.

Les études biophysiques réalisées sur la région P_{NTD} de RV et VSV ont permis de déterminer que cette région de la protéine, fortement conservée et prédite structurée (cf. Chap.II, ARTICLE I), existe en solution sous la forme d'un ensemble de conformères compactes et étendus, et posséde des quantités significatives d'hélices α transitoires (Figure 46A) (cf. Chap. III, ARTICLES II&III). Les résultats obtenus suggèrent fortement que P_{NTD} fonctionne comme un Elément de Reconnaissance Moléculaire (MoRE) (Mohan, Oldfield et al. 2006; Vacic, Oldfield et al. 2007) en se structurant de manière concomitante à sa fixation à N^0.

La structure RMN du domaine P_{CTD} de VSV (Chap. IV, ARTICLE IV), ainsi que la structure cristallographique du domaine P_{CED} de RV (Ivanov, Crepin et al. 2010) sont venues s'ajouter aux structures des domaines P_{CED} du VSV et P_{CTD} de RV, permettant ainsi la modélisation atomique théorique de la phosphoprotéine entière par un ensemble de structures partiellement désordonnées (Figure 46B), qui a pu être comparé au profil SAXS expérimental de manière à obtenir une analyse quantitative du désordre conformationnel de la phosphoprotéine (cf. Chap.V, ARTICLE V). Bien que cette analyse ait été réalisée uniquement chez VSV, la similarité des profils SAXS et l'organisation modulaire dimérique conservée de la protéine suggèrent un désordre similaire pour RV P (cf. Chap.V, Partie III).

Fixation de la phosphoprotéine sur la matrice N-ARN.

L'utilisation d'une procédure d'arrimage croisé flexible (« *cross-docking* ») nous a permis de construire un modèle structural théorique du complexe N-ARN-P_{CTD} de RV à l'aide de diverses contraintes expérimentales ambigües tels que des données SAXS ou encore l'analyse par double hybride de mutants de substitution de P_{CTD} chez le Virus Mokola (cf. Chap.VI, ARTICLE VI). Ces méthodes nous ont permis d'obtenir des informations structurales concernant le complexe N-ARN-P_{CTD} malgré l'échec des tentatives de cristallisation (Francine Gérard, observations non publiées). Bien que la précision de ce type

de modèle obtenu par modélisation moléculaire soit questionnable, les caractéristiques générales du modèle proposé ont été validées à *posteriori* par la structure cristallographique de son homologue chez VSV (Green and Luo 2009), apparue dans la littérature peu de temps après la soumission de nos propres travaux au *Journal of Molecular Biology*.

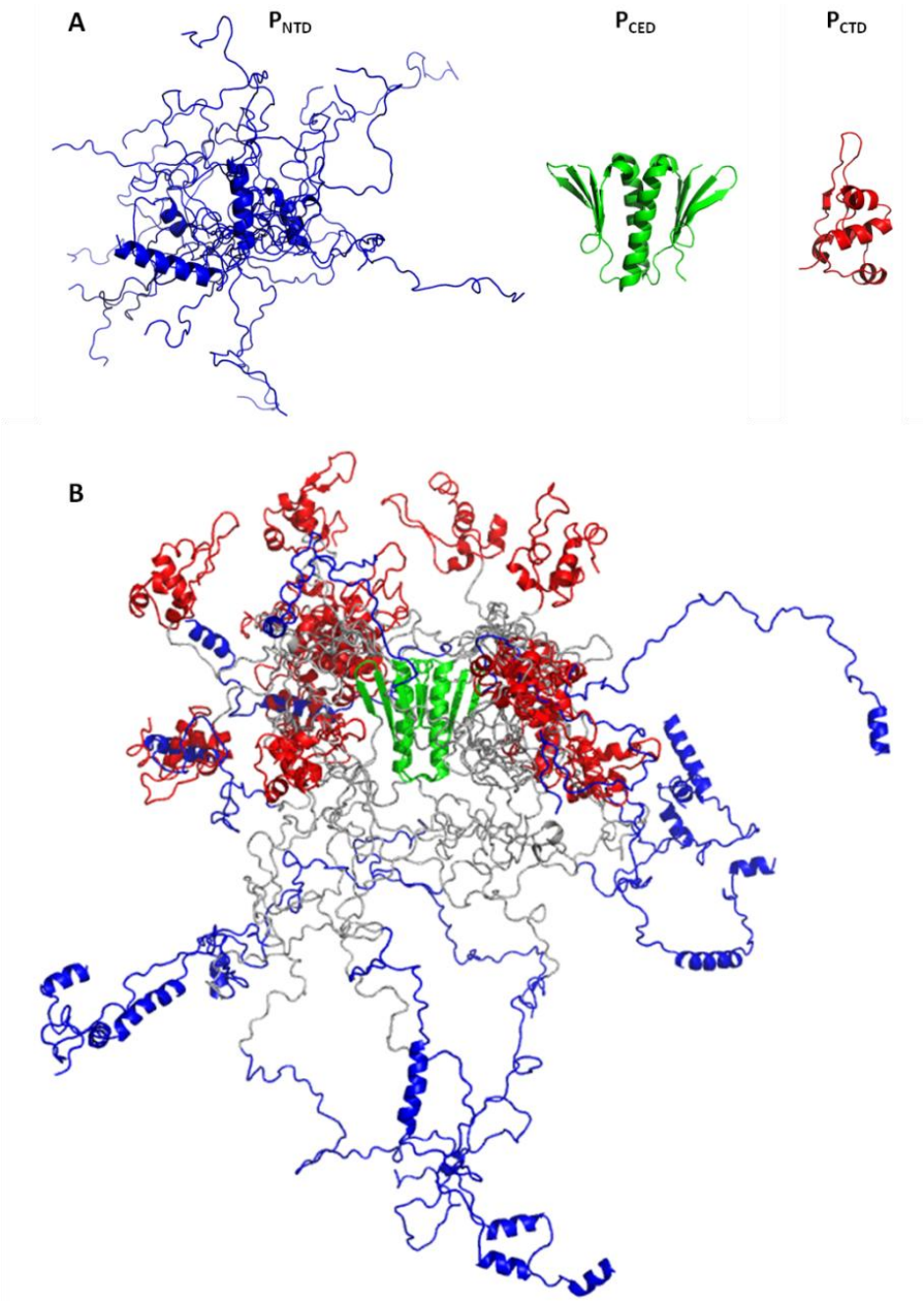


Figure 46: Structure modulaire de la phosphoprotéine illustrée pour VSV. A. De gauche à droite, représentations en cartoon de l'ensemble structural sélectionné sur base du profil SAXS et des données RMN pour P_{NTD} (bleu) (10 structures superposées) (cf. Chap.III, Article II), structure cristallographique du dimère du domaine P_{CED} du VSV (vert) (code pdb 2FQM, (Ding, Green et al. 2006), structure RMN du domaine P_{CTD} du VSV (rouge) (code pdb : 2K47) (cf. Chap.IV, ARTICLE IV). B. Représentation en cartoon de l'ensemble structural sélectionné sur base du profil SAXS pour la phosphoprotéine P_2 (5 modèles superposés) (cf. Chap.V, ARTICLE V). Le code couleur est identique à celui utilisé en A pour P_{NTD}, P_{CED} et P_{CTD}. Les régions IDR_{NTD} et IDR_{CTD} sont représentées en gris. Le modèle souligne la présence d'éléments de structure secondaire α-hélicoïdaux fluctuants dans la région N-terminale, et l'articulation des régions N- et C-terminales autour du domaine central permettant une importante flexibilité de la molécule.

La Figure 47 montre une comparaison des deux structures. Chez VSV comme chez RV, le domaine P_{CTD} interagit simultanément avec deux protomères de nucléoprotéine, et les deux faces opposées de la molécule forment une série de liaisons faibles intermoléculaires avec les boucles C-terminales désordonnées de N, qui tendent à se structurer au contact de P_{CTD}. L'orientation des domaines P_{CTD} est globalement conservée entre les deux virus, avec le N-terminus relié au reste de P pointant en direction de la face interne de N-ARN, suggérant que la région P_{NTD} de la phosphoprotéine est localisée à proximité de la région centrale de fixation à l'ARN de N.

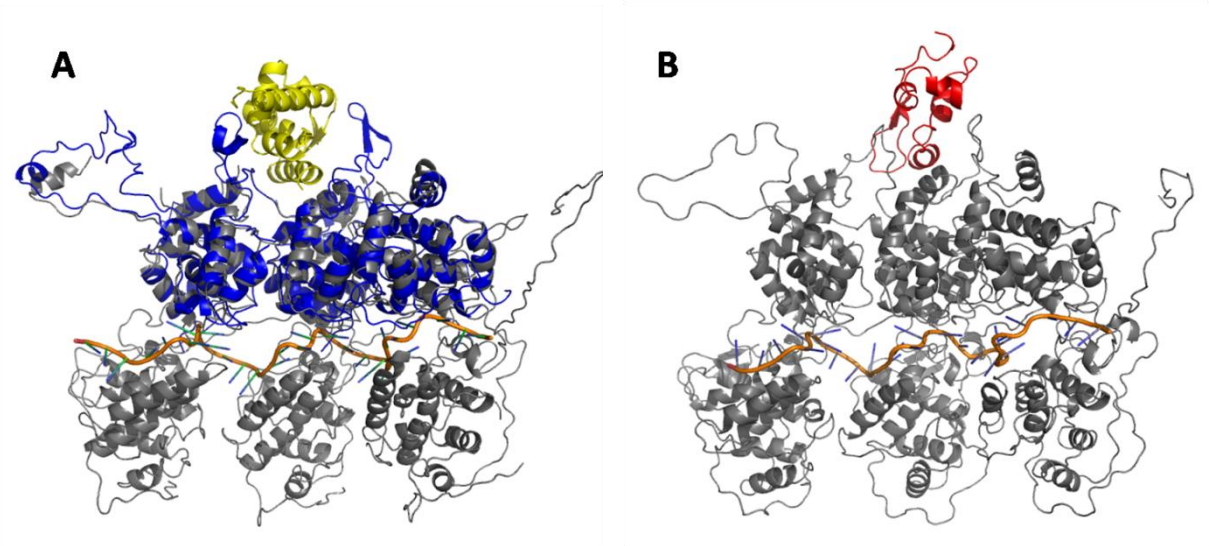


Figure 47: Comparaison entre les structures des complexes N-ARN-P_{CTD} chez RV et VSV. A. Représentation en cartoon du modèle structural du domaine C-terminal de la phosphoprotéine de RV (jaune) associé aux domaines C-terminaux d'un trimère de nucléoprotéines (bleu). Le modèle est superposé sur la structure cristallographique du complexe N-ARN (en gris, code pdb: 2GTT). B. Structure cristallographique du domaine C-terminal de la phosphoprotéine de VSV (rouge) associé aux domaines C-terminaux d'un trimère de N-ARN (gris, code pdb: 3HHZ).

L'interaction entre la phosphoprotéine entière et les complexes N-ARN a été étudiée par SAXS dans le contexte des complexes circulaires N_{10}-ARN-$(P_2)_2$ et N_{11}-ARN-$(P_2)_2$ et les données ont été modélisées par un ensemble de structures afin de prendre en compte le caractère partiellement désordonné de la phosphoprotéine (Chap.VI, Partie III). Les paramètres structuraux globaux dérivés pour cet ensemble de structures sont en bon accord avec l'analyse d'une expérience de variation de contraste entre la phosphoprotéine deutérée et les complexes N-ARN par SANS (Euripedes de Almeida Ribeiro Jr, données non publiées). Cependant, la nature circulaire artefactuelle des complexes N_{10}-ARN et N_{11}-ARN exerce probablement un biais important sur l'occupation de l'espace par P_2 dans ces complexes.

Activité chaperonne de la phosphoprotéine.

L'étude structurale du complexe N^0-P constitue une des clés de la compréhension du mécanisme de réplication virale. En effet, la formation de ce complexe lors de la phase initiale du cycle de réplication permet l'accumulation de nucléoprotéine soluble qui pourra ensuite être utilisée pour l'encapsidation des nouveaux génomes et antigénomes viraux (Figure 49). L'obtention d'information structurale à haute résolution avait jusqu'ici été limitée par la nature transitoire du complexe N^0-P, qui nous a conduits à développer une nouvelle méthode pour la production du complexe (Chap.VII, ARTICLE VII et Partie IV). Nous avons caractérisé le complexe formé entre la région P_{NTD} et la nucléoprotéine soluble N^0 délétée de son sous-domaine d'échange N-terminal *in silico* et *in vitro* par diffraction des rayons X, SAXS, et RMN. Ces expériences ont permis de déterminer pour la première fois une structure à résolution atomique d'un complexe N^0-P d'un Virus de l'ordre des *Mononegavirales*.

Cette structure confirme la stœchiométrie 1 :1 du complexe formé entre la région N-terminale de P et la nucléoprotéine soluble, et met en évidence l'interaction d'une hélice α formée par les résidus 17 à 31 de P_{NTD} avec la cavité de fixation à l'ARN de N, une possibilité qui avait déjà été envisagée auparavant sur bases de données de protéolyse limitée (Mavrakis, Mehouas et al. 2006), et prédite uniquement par modélisation moléculaire (cf. Chap.VII, Partie II).

De manière intéressante, la cristallisation du complexe a été possible grâce à l'utilisation du mutant $N_{\Delta 21}$ de VSV et d'un fragment relativement court de la phosphoprotéine (résidus 1 à 60), permettant ainsi d'éviter la présence de trop longs segments désordonnés. Néanmoins, seuls les résidus 5 à 33 de P_{NTD} sont visibles dans la structure cristallographique, ce qui indique que les résidus 1-4 et 34-60 sont désordonnés dans le cristal, une observation consistente avec la localisation des résidus les plus critiques pour l'interaction entre les positions 11 à 30 (Chen, Ogino et al. 2007). Les données RMN concernant l'interaction suggèrent cependant la fixation de la totalité de la région P_{NTD} sur $N_{\Delta 21}$ et $N_{\Delta 21}{}^{344\text{-}SPSGS\text{-}348}$ en solution, et l'existence d'un échange conformationnel en régime intermédiaire pour la région 1-19 de P_{NTD} (cf. Chap.VII, Partie IV). De plus amples investigations à l'aide de P_{NTD} triplement marqué sont en cours afin de clarifier le rôle des régions 1 à 10 et 35 à 60 dans l'interaction avec N^0 (Filip Yabukarski, données non publiées).

Les données présentées concernant l'interaction de la région P_{NTD} avec les formes mutantes de la nucléoprotéine ($N_{\Delta 21}$ et $N_{\Delta 21}{}^{344\text{-}SPSGS\text{-}348}$ chez VSV, $N_{\Delta 23}$ chez RV) soulignent le rôle critique joué par le bras N-terminal (résidus 1-23 chez RV, 1-21 chez VSV) et la boucle C-terminale de la nucléoprotéine (résidus 351-399 chez RV, 342-374 chez VSV) dans l'assemblage de ces complexes N-ARN. Ces deux régions, qui constituent les deux sous domaines d'échange entre protomères, détiennent les clés de la polymérisation des complexes N-ARN. En effet, l'encapsidation d'ARN non spécifique au sein des complexes circulaires ou linéaires formés par la protéine N sauvage lorsque celle-ci est exprimée seule en cellules d'insectes ou en bactéries est irréversible, comme l'indiquent les profiles d'élution obtenus

par SEC-MALLS-RI en incubant les complexes MBP-N-ARN du VSV avec P_{NTD} en présence ou en absence de RNAseA (Leyrat et Yabukarski, observations non publiées).

Cependant, la déstabilisation du complexe induite par la délétion du domaine d'échange N-terminal rend les phénomènes d'encapsidation et d'oligomérisation réversibles en présence de P_{NTD} et entraîne la libération d'ARN et/ou de protomères de nucléoprotéine à partir des complexes N-ARN initialement présents (cf. Chap.VII, ARTICLE VII). Ce résultat implique que la réversibilité du mécanisme d'encapsidation est lié à l'exposition à la surface du complexe N-ARN du domaine d'échange C-terminal, qui peut ainsi être reconnu par P_{NTD}. Ce phénomène de reconnaissance moléculaire est inhibé dans la forme sauvage de la nucléoprotéine par le masquage de cette surface d'interaction par le domaine d'échange N-terminal. Cette interprétation mécanistique est validée par l'observation, dans la structure cristallographique du complexe $N_{\Delta 21}^{0}$-P_{NTD}, des résidus 5 à 15 de P_{NTD} adoptant une conformation qui mime l'interaction entre les domaines d'échange N- et C-terminaux de la nucléoprotéine observée dans les complexes N-ARN natifs. Le rôle clé du domaine d'échange C-terminal est par ailleurs confirmé par la perte d'affinité relative pour P_{NTD} observée chez le mutant $N_{\Delta 21}^{344\text{-}SPSGS\text{-}348}$, et par les travaux de modélisation structurale du complexe, *ab initio* ou sur base des données cristallographiques et RMN, qui indiquent tous deux une interaction entre les résidus 1 à 12 de P_{NTD} et la boucle C-terminale de N, que celle-ci adopte ou non la conformation α-hélicoïdale observée sous forme résiduelle par RMN en solution pour le P_{NTD} isolé (Chap.III, ARTICLE II).

Le complexes VSV $N_{\Delta 21}^{0}$-P_2 et RV $N_{\Delta 23}^{0}$-P_2 ont été produits respectivement par échange avec les formes VSV $N_{\Delta 21}^{0}$-P_{NTD} et RV $N_{\Delta 23}^{0}$-P_{NTD}. Les complexes ainsi obtenus ont été caractérisés par SEC-MALLS-RI et SAXS (cf. Chap.VII, Partie IV pour RV $N_{\Delta 23}^{0}$-P_2, et données non présentées pour VSV $N_{\Delta 21}^{0}$-P_2). Les masses moléculaires mesurées par MALLS confirment la stoechiométrie 1 : 2 du complexe, en accord avec les données de la littérature disponibles pour RV (Mavrakis, Iseni et al. 2003). L'augmentation des dimensions du complexe par rapport à la phosphoprotéine isolée est très faible, ce qui suggère que l'espace conformationnel de P pourrait être contraint par la fixation à N^0. Une baisse importante de la flexibilité de P pourrait notamment provenir de la fixation additionnelle d'un des deux domaines P_{CTD} sur N^0, comme le suggèrent certaines données préliminaires (Cédric Leyrat et Filip Yabukarski, données non publiées).

La modélisation du complexe N^0-P à l'aide d'un ensemble structural atomique permettrait de comprendre de quelle manière et dans quelles proportions la fixation d'une molécule de N^0 est susceptible de modifier l'espace conformationnel et la flexibilité de la phosphoprotéine, voire d'expliquer l'absence de formation de complexes $(N^0)_2$-P_2 dans les conditions expérimentales. Ces données pourraient ainsi être directement comparées avec l'information disponible pour la phosphoprotéine isolée.

Une autre question intéressante concerne la régulation de la force des interactions entre N^0, P et L. En effet, la phosphoprotéine possèdent plusieurs sites de phosphorylation localisés au niveau des résidus 63 et 64 chez RV et 60, 62 et 64 chez VSV (Takacs, Barik et al. 1992; Hwang, Englund et al. 1999; Gupta, Blondel et al. 2000). Cette région charnière

correspond à la limite entre les sites de fixation pour N^0 et L, et la phosphorylation de P pourrait moduler l'affinité pour l'un, l'autre, ou les deux partenaires. En admettant une conformation triangulaire pour VSV P_{NTD} dans sa forme liée à N^0 en solution, globalement similaire à celle prédite par la modélisation *ab initio* (cf. Chap.VII, Partie II), les sites de phosphorylation se retrouveraient à proximité de l'hélice N-terminale de P_{NTD} et de la boucle C-terminale de N dans le complexe N^0 (cf. Chap.VII, Partie II & III). On peut alors spéculer que la phosphorylation au niveau de ces sites puisse conduire à une stabilisation ou à une destabilisation du complexe N^0-P. Par ailleurs, il est connu que l'activité transcriptionnelle du virus nécessite la phosphorylation de la région N-terminale de P (Barik and Banerjee 1992; Barik and Banerjee 1992; Takacs, Barik et al. 1992) et semble dépendre de son caractère acide (Chattopadhyay and Banerjee 1987; Chattopadhyay and Banerjee 1988), qui module probablement l'interaction avec L. Au contraire, la phosphorylation de la région N-terminale ne semble pas être requise pour la réplication (Pattnaik, Hwang et al. 1997).

La particularité des interactions entre les protéines N et P réside dans leur dualité et l'intime connexion existant entre les deux formes de complexes N-P. Les complexes N^0-P sont en effet utilisés pour former les complexes N-ARN-P, comme cela est illustré par la Figure 48. Il serait ainsi extrêmement intéressant d'étudier l'encapsidation réversible de l'ARN à partir des complexes N^0-P_2 et N^0-P_{NTD} isolés.

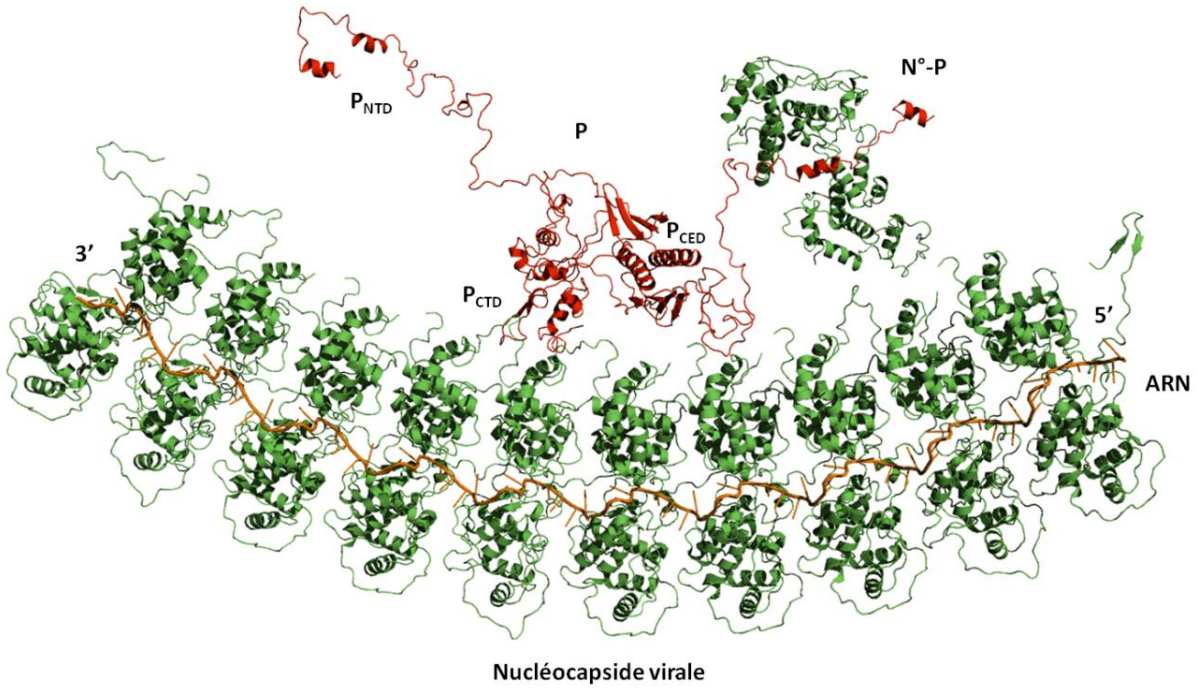


Figure 48 : Alignement entre les structures de N-ARN (Green, Zhang et al. 2006; Ge, Tsao et al. 2010)**, N-ARN-P_{CTD}** (Green and Luo 2009)**, P_2 (Chap.V) et N^0-P_{NTD} (Chap.VII) chez le Virus de la Stomatite Vésiculaire illustrant la dualité des interactions entre les protéines N et P.** La phosphoprotéine est représentée en cartoons rouges, alors que la nucléoprotéine et l'ARN sont colorés en vert et en orange, respectivement. La localisation de la région P_{NTD} et des domaines P_{CED} et P_{CTD} de la phosphoprotéine est indiquée. La conformation de la phosphoprotéine dimérique P_2, ainsi que la position relative des complexes N-ARN et N^0-P sont arbitraires.

Implications pour les mécanismes de la transcription et de la réplication virale chez les Rhabdoviridae

Les mécanismes de la transcription et de la réplication demeurent mal compris, et ce principalement en raison de l'absence de données structurales concernant la polymérase, un problème qui fait actuellement l'objet de nombreuses études. Le travail présenté ici permet néanmoins d'étendre notre description structurale du complexe de réplication des *Rhabdoviridae* en précisant la nature des interactions entre N et P (Chap.VI et VII). Le mécanisme qui semble le plus probable à l'heure actuelle implique la fixation de molécules P_2 à intervalles réguliers sur la nucléocapside virale, permettant l'attachement de la polymérase L au P_{NTD}. La polymérase peut ensuite transcrire l'ARN en se déplaçant sur la matrice N-ARN et en fixant les phosphoprotéines disposées le long de la nucléocapside (Figure 49), en tirant ainsi profit de la nature flexible et allongée de P_2. Ce modèle est appuyé principalement par les valeurs d'affinité élevées mesurées pour l'interaction entre P et N-ARN (Chap.VI, ARTICLE VI ; Euripedes de Almeida Ribeiro Jr, données non publiées) et par l'observation que le domaine de dimérisation n'est pas requis pour la transcription (Jacob, Real et al. 2001). Par ailleurs, l'affinité de la région N-terminale de P pour la polymérase semble relativement faible, cette région pouvant être remplacée par des séquences acides similaires (Chattopadhyay and Banerjee 1987; Chattopadhyay and Banerjee 1988). Ces observations sont en outre consistantes avec la capacité de P_2 à utiliser ses bras désordonnés pour pêcher (« fly casting ») ses partenaires tels que les protéines L et N^0 à des distances importantes, pouvant dépassant les 15 nm (Chap.V). Ensemble, ces données suggèrent un déplacement de L sur le complexe N-ARN-P, dans lequel la phosphoprotéine ne se dissocie pas de la matrice N-ARN durant la synthèse d'ARN, mais fixe la protéine L au fur-et-à-mesure de ses déplacement sur la matrice. Ce mécanisme permet à la phosphoprotéine d'assurer un positionnement correct de la polymérase sur la matrice N-ARN et d'acheminer les complexes N^0-P au site d'encapsidation lors de la réplication. Ce modèle diffère du modèle de type « cartwheeling » proposé pour les *Paramyxoviridae* qui implique un mouvement de P sur la matrice N-ARN *via* une rotation autour de l'axe d'ordre 4 du domaine central et l'association et la dissociation continuelle des domaines PX sur le complexe N-ARN (Curran and Kolakofsky 1999; Kolakofsky, Le Mercier et al. 2004).

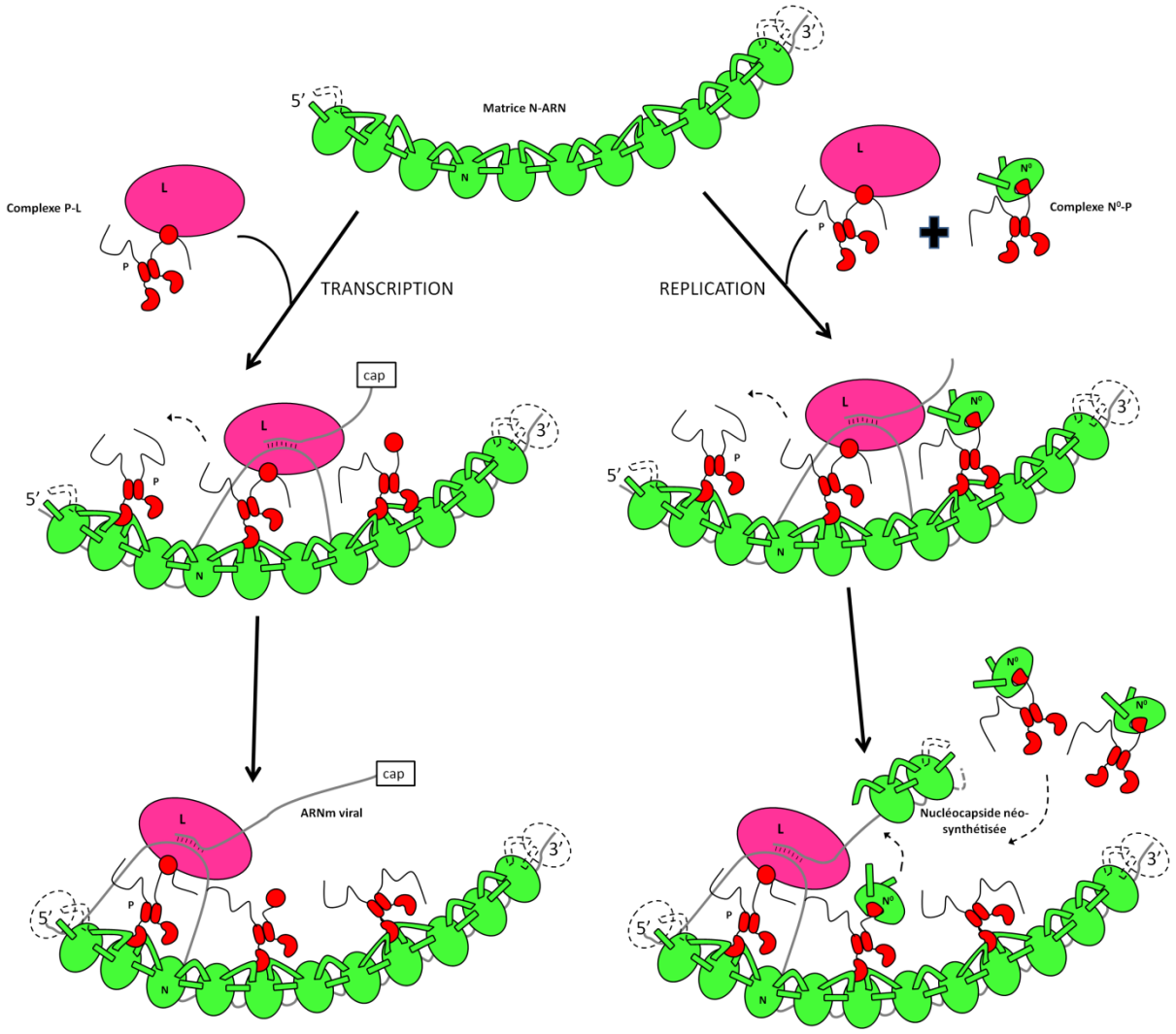


Figure 49: Représentation Schématique des mécanismes de transcription et de réplication chez les *Rhabdoviridae*. Au cours de la synthèse de l'ARN viral, la polymérase doit se déplacer le long de la nucléocapside et accéder à l'ARN séquestré dans la cavité centrale de la nucléoprotéine. P fournit le lien physique par lequel la polymérase s'associe à sa matrice. Chez RV, le domaine de dimérisation de P n'est pas requis pour la transcription et la phosphoprotéine, ce qui suggère que P n'utilise pas le mécanisme de « cartwheeling » pour se mouvoir sur la matrice N-ARN. Etant donné la forte affinité de P_{CTD} pour N-RNA, on peut supposer que les dimères P_2 sont fixés à intervalles régulier le long des nucléocapsides virales. Au cours de la transcription, la polymérase se lie à la région N-terminale de P et transcrit le brin d'ARN viral. Comme P est allongée et flexible, le bras N-terminal de la molécule P_2 adjacente peut s'étendre pour fixer la polymérase, qui est ensuite transférée en aval. La polymérase peut ainsi se déplacer le long de la matrice N-RNA en sautant d'une molécule P_2 à l'autre. Au cours de la réplication, la polymérase peut également se mouvoir selon un mode similaire, mais les génomes et antigénomes nouvellement synthétisés sont cette fois encapsidés par N. Les molécules de nucléoprotéines nouvellement produites sont présentes sous la forme de complexes N^0-P solubles, qui peuvent également se fixer au complexe N-RNA *via* leurs domaines P_{CTD}. N^0 peut ensuite être acheminé jusqu'au site de synthèse d'ARN pour produire de nouveaux complexes N-RNA.

REFERENCES BIBILIOGRAPHIQUES

gcq#236: "Wait a Minute, aren't You.... ? (gunshots) Yeah."

(Bodycount)

Abraham, G. and A. K. Banerjee (1976). "Sequential transcription of the genes of vesicular stomatitis virus." Proc Natl Acad Sci U S A **73**(5): 1504-8.
Abraham, G., D. P. Rhodes, et al. (1975). "The 5' terminal structure of the methylated mRNA synthesized in vitro by vesicular stomatitis virus." Cell **5**(1): 51-8.
Adcock, S. A. and J. A. McCammon (2006). "Molecular dynamics: survey of methods for simulating the activity of proteins." Chem Rev **106**(5): 1589-615.
Ago, H., T. Adachi, et al. (1999). "Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus." Structure **7**(11): 1417-26.
Ahmed, M. and D. S. Lyles (1998). "Effect of vesicular stomatitis virus matrix protein on transcription directed by host RNA polymerases I, II, and III." Journal of Virology **72**(10): 8413-8413.
Albertini, A. A., C. R. Clapier, et al. (2006). "Isolation and crystallization of a unique size category of recombinant Rabies virus Nucleoprotein-RNA rings." J Struct Biol **158**: 129-133.
Albertini, A. A., G. Schoehn, et al. (2008). "Structural aspects of rabies virus replication." Cell Mol Life Sci **65**(2): 282-94.
Albertini, A. A., A. K. Wernimont, et al. (2006). "Crystal structure of the rabies virus nucleoprotein-RNA complex." Science **313**(5785): 360-3.
Albertini, A. A. V., G. Schoehn, et al. (2005). "Structures impliquées dans la réplication et la transcription des virus à ARN non segmenté de sens négatif." Virologie **9**(2): 83-92.
Albertini, A. A. V., A. K. Wernimont, et al. (2006). "Crystal structure of the rabies virus nucleoprotein-RNA complex." Science **313**(5785): 360-360.
Alvarez, L., R. Fajardo, et al. (1994). "Partial recovery from rabies in a nine-year-old boy." The Pediatric infectious disease journal **12**(12): 1154-1154.
Ammar, E. D. and S. A. Hogenhout (2008). "A neurotropic route for Maize mosaic virus (Rhabdoviridae) in its planthopper vector Peregrinus maidis." Virus Research **131**(1): 77-85.
Area, E., J. Martin-Benito, et al. (2004). "3D structure of the influenza virus polymerase complex: localization of subunit domains." Proc Natl Acad Sci U S A **101**(1): 308-13.
Arnheiter, H., N. L. Davis, et al. (1985). "Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis." Cell **41**(1): 259-67.
Arnold, K., L. Bordoli, et al. (2006). "The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling." Bioinformatics **22**(2): 195-201.
Assenberg, R., O. Delmas, et al. (2010). "Genomics and structure/function studies of Rhabdoviridae proteins involved in replication and transcription." Antiviral Res **87**(2): 149-61.
Assenberg, R., O. Delmas, et al. (2009). "The Structure of the N-RNA Binding Domain of the Mokola virus Phosphoprotein." J Virol: Nov 11. [Epub ahead of print]PMID: 19906936 [PubMed - as supplied by publisher].
Astic, L., D. Saucier, et al. (1993). "The CVS strain of rabies virus as transneuronal tracer in the olfactory system of mice." Brain research **619**(1-2): 146-156.
Atanasiu, P., P. Lepine, et al. (1963). "Etude morphologique du virus rabique des rues en culture de tissu." Compt. Rend **256**: 3219-3221.
Audie, J. (2009). "Continued development of an empirical function for predicting and rationalizing protein-protein binding affinities." Biophys Chem **143**(3): 139-44.
Audie, J. and S. Scarlata (2007). "A novel empirical free energy function that explains and predicts protein-protein binding affinities." Biophys Chem **129**(2-3): 198-211.
Babes, V. (1912). "Traite de la rage."
Badrane, H., C. Bahloul, et al. (2001). "Evidence of two Lyssavirus phylogroups with distinct pathogenicity and immunogenicity." J Virol **75**(7): 3268-76.
Badrane, H. and N. Tordo (2001). "Host switching in Lyssavirus history from the Chiroptera to the Carnivora orders." J Virol **75**(17): 8096-104.
Baer, G. M., J. H. Shaddock, et al. (1990). "Rabies susceptibility and acetylcholine receptor." Lancet **335**(8690): 664-5.
Bahri, F., A. Letaief, et al. (1996). "Neurological complications in adults following rabies vaccine prepared from animal brains." Presse mÃŠdicale (Paris, France: 1983) **25**(10): 491-491.
Baldwin, R. L. and G. D. Rose (1999). "Is protein folding hierarchic? I. Local structure and peptide folding." Trends Biochem Sci **24**(1): 26-33.
Ball, L. A. (2007). "Virus replication strategies." Fields Virology: 131-139.
Ball, L. A. and C. N. White (1976). "Order of transcription of genes of vesicular stomatitis virus." Proc Natl Acad Sci U S A **73**(2): 442-6.
Baltimore, D., A. S. Huang, et al. (1970). "Ribonucleic acid synthesis of vesicular stomatitis virus, II. An RNA polymerase in the virion." Proc Natl Acad Sci U S A **66**(2): 572-6.

Bandyopadhyay, A., K. Kopperud, et al. (2010). "An integrated protein localization and interaction map for Potato yellow dwarf virus, type species of the genus Nucleorhabdovirus." Virology.
Banerjee, A. D., G. Abraham, et al. (1977). "Vesicular stomatitis virus: mode of transcription." J Gen Virol **34**(1): 1-8.
Banerjee, A. K. (1987). "Transcription and replication of rhabdoviruses." Microbiol Rev **51**(1): 66-87.
Banerjee, A. K. (1987). "The transcription complex of vesicular stomatitis virus." Cell **48**(3): 363-4.
Banerjee, A. K., G. Abraham, et al. (1977). "Vesicular stomatitis virus: mode of transcription." Journal of General Virology **34**(1): 1-1.
Banerjee, A. K. and S. Barik (1992). "Gene expression of vesicular stomatitis virus genome RNA." Virology **188**(2): 417-28.
Banerjee, A. K., S. A. Moyer, et al. (1974). "Studies on the in vitro adenylation of RNA by vesicular stomatitis virus." Virology **61**(2): 547-58.
Barge, A., J. Gagnon, et al. (1996). "Rod-like shape of vesicular stomatitis virus matrix protein." Virology **219**(2): 465-70.
Barge, A., Y. Gaudin, et al. (1993). "Vesicular stomatitis virus M protein may be inside the ribonucleocapsid coil." J Virol **67**(12): 7246-53.
Barik, S. and A. K. Banerjee (1992). "Phosphorylation by cellular casein kinase II is essential for transcriptional activity of vesicular stomatitis virus phosphoprotein P." Proc Natl Acad Sci U S A **89**(14): 6570-4.
Barik, S. and A. K. Banerjee (1992). "Sequential phosphorylation of the phosphoprotein of vesicular stomatitis virus by cellular and viral protein kinases is essential for transcription activation." Journal of Virology **66**(2): 1109-1109.
Barik, S. and A. K. Banerjee (1992). "Sequential phosphorylation of the phosphoprotein of vesicular stomatitis virus by cellular and viral protein kinases is essential for transcription activation." J Virol **66**(2): 1109-18.
Barr, J. N., S. P. Whelan, et al. (1997). "cis-Acting signals involved in termination of vesicular stomatitis virus mRNA synthesis include the conserved AUAC and the U7 signal for polyadenylation." Journal of Virology **71**(11): 8718-8718.
Barr, J. N., S. P. Whelan, et al. (2002). "Transcriptional control of the RNA-dependent RNA polymerase of vesicular stomatitis virus." Biochim Biophys Acta **1577**(2): 337-53.
Baskakov, I. and D. W. Bolen (1998). "Forcing thermodynamically unfolded proteins to fold." J Biol Chem **273**(9): 4831-4.
Baskakov, I. V., R. Kumar, et al. (1999). "Trimethylamine N-oxide-induced cooperative folding of an intrinsically unfolded transcription-activating fragment of human glucocorticoid receptor." J Biol Chem **274**(16): 10693-6.
Baudin, F., C. Bach, et al. (1994). "Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent." Embo J **13**(13): 3158-65.
Beijerinck, M. W. (1898). "Ueber ein contagium vivum fluidum als Ursache der Fleckenkrankheit der Tabaksblatter." Verh. Kon. Akad. Wetensch **5**: 3-21.
Benmansour, A., H. Leblois, et al. (1991). "Antigenicity of rabies virus glycoprotein." Journal of Virology **65**(8): 4198-4198.
Berendsen, H. J. C., J. R. Grigera, et al. (1987). "The missing term in effective pair potentials." Journal of Physical Chemistry **91**(24): 6269-6271.
Berg, M., B. Hjertner, et al. (1992). "The P gene of the procine paramyxovirus LPMV encodes three possible polypeptides P, V and C: the P protein mRNA is edited." Journal of General Virology **73**(5): 1195-1195.
Bergmann, J. E., K. T. Tokuyasu, et al. (1981). "Passage of an integral membrane protein, the vesicular stomatitis virus glycoprotein, through the Golgi apparatus en route to the plasma membrane." Proceedings of the National Academy of Sciences of the United States of America **78**(3): 1746-1746.
Bernado, P. (2010). "Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering." Eur Biophys J **39**(5): 769-80.
Bernado, P. and M. Blackledge (2009). "A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering." Biophys J **97**(10): 2839-45.
Bernado, P., L. Blanchard, et al. (2005). "A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering." Proc Natl Acad Sci U S A **102**(47): 17002-7.
Bernado, P., E. Mylonas, et al. (2007). "Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering." J Am Chem Soc.
Bernado, P., E. Mylonas, et al. (2007). "Structural characterization of flexible proteins using small-angle X-ray scattering." J Am Chem Soc **129**(17): 5656-64.
Bernardi, R. and P. P. Pandolfi (2007). "Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies." Nat Rev Mol Cell Biol **8**(12): 1006-16.

Bernauer, J., J. Aze, et al. (2007). "A new protein-protein docking scoring function based on interface residue properties." Bioinformatics **23**(5): 555-62.
Bhella, D., A. Ralph, et al. (2002). "Significant differences in nucleocapsid morphology within the Paramyxoviridae." Journal of General Virology **83**(8): 1831-1831.
Bhella, D., A. Ralph, et al. (2004). "Conformational flexibility in recombinant measles virus nucleocapsids visualised by cryo-negative stain electron microscopy and real-space helical reconstruction." Journal of Molecular Biology **340**(2): 319-331.
Bishop, D. H. and P. Roy (1972). "Dissociation of vesicular stomatitis virus and relation of the virion proteins to the viral transcriptase." J Virol **10**(2): 234-43.
Black, B. L. and D. S. Lyles (1992). "Vesicular stomatitis virus matrix protein inhibits host cell-directed transcription of target genes in vivo." Journal of Virology **66**(7): 4058-4058.
Blanchard, L., N. Tarbouriech, et al. (2004). "Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution." Virology **319**(2): 201-11.
Bleck, T. P. and C. E. Rupprecht (2000). "Rabies virus." Mandell GL, Bennett JE, Dolin R. Principles and Practice of Infectious Diseases. 5th ed. Churchill Livingstone **1811**(9).
Blondel, D., S. Kheddache, et al. (2010). "Resistance to Rabies viral infection conferred by PMLIV isoform." J Virol.
Blondel, D., T. Regad, et al. (2002). "Rabies virus P and small P products interact directly with PML and reorganize PML nuclear bodies." Oncogene **21**(52): 7957-7970.
Blondel, D., T. Regad, et al. (2002). "Rabies virus P and small P products interact directly with PML and reorganize PML nuclear bodies." Oncogene **21**(52): 7957-70.
Blumberg, B. M., C. Giorgi, et al. (1983). "N protein of vesicular stomatitis virus selectively encapsidates leader RNA in vitro." Cell **32**(2): 559-67.
Blumberg, B. M., C. Giorgi, et al. (1984). "Preparation and analysis of the nucleocapsid proteins of vesicular stomatitis virus and sendai virus, and analysis of the sendai virus leader-NP gene region." J Gen Virol **65 (Pt 4)**: 769-79.
Blumberg, B. M., M. Leppert, et al. (1981). "Interaction of VSV leader RNA and nucleocapsid protein may control VSV genome replication." Cell **23**(3): 837-45.
Bogerd, H. P., R. A. Fridell, et al. (1996). "Protein sequence requirements for function of the human T-cell leukemia virus type 1 Rex nuclear export signal delineated by a novel in vivo randomization-selection assay." Mol Cell Biol **16**(8): 4207-14.
Bolen, D. W. and I. V. Baskakov (2001). "The osmophobic effect: natural selection of a thermodynamic force in protein folding." J Mol Biol **310**(5): 955-63.
Bourhis, J. M., K. Johansson, et al. (2004). "The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner." Virus Res **99**(2): 157-67.
Bourhy, H., B. Kissi, et al. (1993). "Molecular diversity of the Lyssavirus genus." Virology **194**(1): 70-81.
Bourhy, H., N. Tordo, et al. (1989). "Complete cloning and molecular organization of a rabies-related virus, Mokola virus." J Gen Virol **70 (Pt 8)**: 2063-74.
Broughan, J. H. and W. H. Wunner (1995). "Characterization of protein involvement in rabies virus binding to BHK-21 cells." Arch Virol **140**(1): 75-93.
Brun, G., X. Bao, et al. (1995). "The relationship of Piry virus to other vesiculoviruses: a re-evaluation based on the glycoprotein gene sequence." Intervirology **38**(5): 274-282.
Brzozka, K., S. Finke, et al. (2005). "Identification of the rabies virus alpha/beta interferon antagonist: phosphoprotein P interferes with phosphorylation of interferon regulatory factor 3." J Virol **79**(12): 7673-81.
Brzozka, K., S. Finke, et al. (2006). "Inhibition of interferon signaling by rabies virus phosphoprotein P: activation-dependent binding of STAT1 and STAT2." J Virol **80**(6): 2675-83.
Buchen-Osmond, C. (2003). "The universal virus database ICTVdB." Computing in Science & Engineering **5**(3): 16-25.
Büchen-Osmond, C. (2003). "Taxonomy and classification of viruses." Manual of clinical microbiology **2**: 1217–1226.
Bukreyev, A., M. H. Skiadopoulos, et al. (2006). "Nonsegmented negative-strand viruses as vaccine vectors." J Virol **80**(21): 10293-306.
Burrage, T. G., G. H. Tignor, et al. (1985). "Rabies virus binding at neuromuscular junctions." Virus Res **2**(3): 273-89.
Bussi, G., D. Donadio, et al. (2007). "Canonical sampling through velocity rescaling " J. Chem. Phys. **126**: 014101.
Byrappa, S., Y. B. Pan, et al. (1996). "Sendai virus P protein is constitutively phosphorylated at serine249: high phosphorylation potential of the P protein." Virology **216**(1): 228-34.

Calain, P. and L. Roux (1993). "The rule of six, a basic feature for efficient replication of Sendai virus defective interfering RNA." J Virol **67**(8): 4822-30.
Camacho, C. J. and C. Zhang (2005). "FastContact: rapid estimate of contact and binding free energies." Bioinformatics **21**(10): 2534-2534.
Canter, D. M. and J. Perrault (1996). "Stabilization of vesicular stomatitis virus L polymerase protein by P protein binding: a small deletion in the C-terminal domain of L abrogates binding." Virology **219**(2): 376-86.
Cantlon, J. D., P. W. Gordy, et al. (2000). "Immune responses in mice, cattle and horses to a DNA vaccine for vesicular stomatitis." Vaccine **18**(22): 2368-2374.
Carneiro, F. A., M. L. Bianconi, et al. (2002). "Membrane recognition by vesicular stomatitis virus involves enthalpy-driven protein-lipid interactions." Journal of Virology **76**(8): 3756-3756.
Carneiro, F. A., A. S. Ferradosa, et al. (2001). "Low pH-induced conformational changes in vesicular stomatitis virus glycoprotein involve dramatic structure reorganization." Journal of Biological Chemistry **276**(1): 62-62.
Carneiro, F. A., P. A. Lapido-Loureiro, et al. (2006). "Probing the interaction between vesicular stomatitis virus and phosphatidylserine." European Biophysics Journal **35**(2): 145-154.
Carneiro, F. A., F. Stauffer, et al. (2003). "Membrane fusion induced by vesicular stomatitis virus depends on histidine protonation." J Biol Chem **278**(16): 13789-94.
Castel, G., M. Chteoui, et al. (2009). "Peptides that mimic the amino-terminal end of the rabies virus phosphoprotein have antiviral activity." J Virol **83**(20): 10808-20.
Celinski, S. A. and J. M. Scholtz (2002). "Osmolyte effects on helix formation in peptides and the stability of coiled-coils." Protein Sci **11**(8): 2048-51.
Cevik, B., S. Smallwood, et al. (2007). "Two N-terminal regions of the Sendai virus L RNA polymerase protein participate in oligomerization." Virology **363**(1): 189-97.
Chamberlin, M. and J. Ring (1973). "Characterization of T7-specific ribonucleic acid polymerase." Journal of Biological Chemistry **248**(6): 2235-2235.
Champ, P. C. and C. J. Camacho (2007). "FastContact: a free energy scoring tool for protein-protein complex structures." Nucleic acids research **35**(Web Server issue): W556-W556.
Champ, P. C. and C. J. Camacho (2007). "FastContact: a free energy scoring tool for protein-protein complex structures." Nucleic Acids Res **35**(Web Server issue): W556-60.
Chattopadhyay, D. and A. K. Banerjee (1987). "Phosphorylation within a specific domain of the phosphoprotein of vesicular stomatitis virus regulates transcription in vitro." Cell **49**(3): 407-14.
Chattopadhyay, D. and A. K. Banerjee (1987). "Two separate domains within vesicular stomatitis virus phosphoprotein support transcription when added in trans." Proc Natl Acad Sci U S A **84**(24): 8932-6.
Chattopadhyay, D. and A. K. Banerjee (1988). "NH2-terminal acidic region of the phosphoprotein of vesicular stomatitis virus can be functionally replaced by tubulin." Proc Natl Acad Sci U S A **85**(21): 7977-81.
Chelbi-Alix, M. K., A. Vidy, et al. (2006). "Rabies viral mechanisms to escape the IFN system: the viral protein P interferes with IRF-3, Stat1, and PML nuclear bodies." Journal of Interferon & Cytokine Research **26**(5): 271-280.
Chelbi-Alix, M. K., A. Vidy, et al. (2006). "Rabies viral mechanisms to escape the IFN system: the viral protein P interferes with IRF-3, Stat1, and PML nuclear bodies." J Interferon Cytokine Res **26**(5): 271-80.
Chen, J. L., T. Das, et al. (1997). "Phosphorylated states of vesicular stomatitis virus P protein in vitro and in vivo." Virology **228**(2): 200-12.
Chen, M., T. Ogino, et al. (2007). "Interaction of vesicular stomatitis virus P and N proteins: Identification of two overlapping domains at the N-terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA." J Virol **81**: 13478-13485.
Chen, R., L. Li, et al. (2003). "ZDOCK: an initial-stage protein-docking algorithm." Proteins **52**(1): 80-7.
Chen, Z., T. J. Green, et al. (2004). "Visualizing the RNA molecule in the bacterially expressed vesicular stomatitis virus nucleoprotein-RNA complex." Structure **12**(2): 227-35.
Chenik, M., K. Chebli, et al. (1995). "Translation initiation at alternate in-frame AUG codons in the rabies virus phosphoprotein mRNA is mediated by a ribosomal leaky scanning mechanism." J Virol **69**(2): 707-12.
Chenik, M., K. Chebli, et al. (1994). "In vivo interaction of rabies virus phosphoprotein (P) and nucleoprotein (N): existence of two N-binding sites on P protein." J Gen Virol **75** 2889-96.
Chenik, M., M. Schnell, et al. (1998). "Mapping the interacting domains between the rabies virus polymerase and phosphoprotein." J Virol **72**(3): 1925-30.
Childs, K. S., J. Andrejeva, et al. (2009). "Mechanism of mda-5 Inhibition by paramyxovirus V proteins." J Virol **83**(3): 1465-73.
Chong, L. D. and J. K. Rose (1993). "Membrane association of functional vesicular stomatitis virus matrix protein in vivo." Journal of Virology **67**(1): 407-407.

Choudhary, S. K., A. G. Malur, et al. (2002). "Characterization of the oligomerization domain of the phosphoprotein of human parainfluenza virus type 3." Virology **302**(2): 373-82.
Chow, T. L., F. H. Chow, et al. (1954). "Morphology of vesicular stomatitis virus." J Bacteriol **68**(6): 724-6.
Chua, K. B., W. J. Bellini, et al. (2000). "Nipah virus: a recently emergent deadly paramyxovirus." Science **288**(5470): 1432-5.
Chua, K. B., S. K. Lam, et al. (2000). "High mortality in Nipah encephalitis is associated with presence of virus in cerebrospinal fluid." Ann Neurol **48**(5): 802-5.
Cisneros, F. M., C. Tsai, et al. (2008). "Nuclear import of Maize fine streak virus proteins in Drosophila S 2 cells." Phytopathology **98**(6).
Clague, M. J., C. Schoch, et al. (1990). "Gating kinetics of pH-activated membrane fusion of vesicular stomatitis virus with cells: stopped-flow measurements by dequenching of octadecylrhodamine fluorescence." Biochemistry **29**(5): 1303-8.
Cleaveland, S., E. M. Fevre, et al. (2002). "Estimating human rabies mortality in the United Republic of Tanzania from dog bite injuries." Bull World Health Organ **80**(4): 304-10.
Cliquet, F. and M. Aubert (2004). "Elimination of terrestrial rabies in Western European countries." Developments in biologicals **119**: 185-204.
Colonno, R. J., G. Abraham, et al. (1976). "Blocked and unblocked 5'termini in vesicular stomatitis virus product RNA in vitro: their possible role in mRNA biosynthesis." Prog Nucleic Acid Res Mol Biol **19**: 83-7.
Colonno, R. J. and A. K. Banerjee (1977). "Mapping and initiation studies on the leader RNA of vesicular stomatitis virus." Virology **77**(1): 260-8.
Colonno, R. J. and A. K. Banerjee (1978). "Complete nucleotide sequence of the leader RNA synthesized in vitro by vesicular stomatitis virus." Cell **15**(1): 93-101.
Comer, J. A., R. B. Tesh, et al. (1990). "Vesicular stomatitis virus, New Jersey serotype: replication in and transmission by Lutzomyia shannoni (Diptera: Psychodidae)." The American journal of tropical medicine and hygiene **42**(5): 483-483.
Compans, R. W. and P. W. Choppin (1967). "Isolation and properties of the helical nucleocapsid of the parainfluenza virus SV5." Proceedings of the National Academy of Sciences of the United States of America **57**(4): 949-956.
Condra, J. H. and R. A. Lazzarini (1980). "Replicative RNA synthesis and nucleocapsid assembly in vesicular stomatitis virus-infected permeable cells." J Virol **36**(3): 796-804.
Connor, J. H., M. O. McKenzie, et al. (2006). "Role of residues 121 to 124 of vesicular stomatitis virus matrix protein in virus assembly and virus-host interaction." J Virol **80**(8): 3701-11.
Conti, C., F. Superti, et al. (1986). "Membrane carbohydrate requirement for rabies virus binding to chicken embryo related cells." Intervirology **26**(3): 164-8.
Cotton, W. E. (1927). "Vesicular stomatitis." Vet. Med **22**: 169-175.
Cowtan, K. D. (2006). "The Buccaneer software for automated model building." Acta Crystallographica Section D-Biological Crystallography **62**(9): 1002-1011.
Crimmins, D. L., W. B. Mehard, et al. (1983). "Physical properties of a soluble form of the glycoprotein of vesicular stomatitis virus at neutral and acidic pH." Biochemistry **22**(25): 5790-6.
Cureton, D. K., R. H. Massol, et al. (2009). "Vesicular Stomatitis Virus Enters Cells through Vesicles Incompletely Coated with Clathrin that depend upon actin for internalization."
Curran, J. and D. Kolakofsky (1999). "Replication of paramyxoviruses." Adv Virus Res **54**: 403-22.
Curran, J., J. B. Marq, et al. (1995). "An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication." J Virol **69**(2): 849-55.
Dancho, B., M. O. McKenzie, et al. (2009). "Vesicular stomatitis virus matrix protein mutations that affect association with host membranes and viral nucleocapsids." J Biol Chem **284**(7): 4500-9.
Darnell, S. J., L. LeGault, et al. (2008). "KFC Server: interactive forecasting of protein interaction hot spots." Nucleic acids research **36**(Web Server issue): W265-W265.
Das, S. C., D. Panda, et al. (2009). "Biarsenical labeling of vesicular stomatitis virus encoding tetracysteine-tagged M protein allows dynamic imaging of M protein and virus uncoating in infected cells." Journal of Virology **83**(6): 2611-2611.
Das, S. C. and A. K. Pattnaik (2004). "Phosphorylation of vesicular stomatitis virus phosphoprotein P is indispensable for virus growth." J Virol **78**(12): 6420-30.
Das, S. C. and A. K. Pattnaik (2005). "Role of the hypervariable hinge region of phosphoprotein P of vesicular stomatitis virus in viral RNA synthesis and assembly of infectious virus particles." J Virol **79**(13): 8101-12.
Das, T. and A. K. Banerjee (1993). "Acidic domain of the phosphoprotein (P) of vesicular stomatitis virus differentially interacts with homologous and heterologous nucleocapsid protein (N)." Cell Mol Biol Res **39**(2): 93-100.

Das, T., A. K. Gupta, et al. (1995). "Role of cellular casein kinase II in the function of the phosphoprotein (P) subunit of RNA polymerase of vesicular stomatitis virus." J Biol Chem **270**(41): 24100-7.
Das, T., M. Mathur, et al. (1998). "RNA polymerase of vesicular stomatitis virus specifically associates with translation elongation factor-1 alphabetagamma for its activity." Proc Natl Acad Sci U S A **95**(4): 1449-54.
Das, T., A. K. Pattnaik, et al. (1997). "Basic amino acid residues at the carboxy-terminal eleven amino acid region of the phosphoprotein (P) are required for transcription but not for replication of vesicular stomatitis virus genome RNA." Virology **238**(1): 103-14.
Das, T., A. Schuster, et al. (1995). "Involvement of cellular casein kinase II in the phosphorylation of measles virus P protein: identification of phosphorylation sites." Virology **211**(1): 218-26.
David, A. E. (1973). "Assembly of the vesicular stomatitis virus envelope: Incorporation of viral polypeptides into the host plasma membrane* 1." Journal of Molecular Biology **76**(1): 135-148.
Davis, N. L., H. Arnheiter, et al. (1986). "Vesicular stomatitis virus N and NS proteins form multiple complexes." J Virol **59**(3): 751-4.
Dawson, R., L. Muller, et al. (2003). "The N-terminal domain of p53 is natively unfolded." J Mol Biol **332**(5): 1131-41.
De, B. P. and A. K. Banerjee (1984). "Specific interactions of vesicular stomatitis virus L and NS proteins with heterologous genome ribonucleoprotein template lead to mRNA synthesis in vitro." J Virol **51**(3): 628-34.
De, B. P. and A. K. Banerjee (1985). "Requirements and functions of vesicular stomatitis virus L and NS proteins in the transcription process in vitro." Biochem Biophys Res Commun **126**(1): 40-9.
Delaglio, F., S. Grzesiek, et al. (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." J Biomol NMR **6**(3): 277-93.
DeLano, W. L. (2002). The PyMOL Molecular Graphics System, DeLano Scientific, Palo Alto, CA, USA.
Delhaye, S., S. Paul, et al. (2006). "Neurons produce type I interferon during viral encephalitis." Proc Natl Acad Sci U S A **103**(20): 7835-40.
Delmas, O., E. C. Holmes, et al. (2008). "Genomic diversity and evolution of the lyssaviruses." PLoS One **3**(4): e2057.
Deng, M., J. N. Bragg, et al. (2007). "Role of the sonchus yellow net virus N protein in formation of nuclear viroplasms." Journal of virology **81**(10): 5362-5362.
Derewenda, Z. S. (2004). "Rational protein crystallization by mutational surface engineering." Structure **12**(4): 529-35.
Derewenda, Z. S. and P. G. Vekilov (2006). "Entropy and surface engineering in protein crystallization." Acta Crystallogr D Biol Crystallogr **62**(Pt 1): 116-24.
Desmezieres, E., A. P. Maillard, et al. (2003). "Differential stability and fusion activity of Lyssavirus glycoprotein trimers." Virus Res **91**(2): 181-7.
Dessen, A., V. Volchkov, et al. (2000). "Crystal structure of the matrix protein VP40 from Ebola virus." EMBO J **19**(16): 4228-36.
Dias, A., D. Bouvier, et al. (2009). "The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit." Nature **458**(7240): 914-8.
Dietzschold, B., J. Li, et al. (2008). "Concepts in the pathogenesis of rabies." Future Virol **3**(5): 481-490.
Dilley, K. A., D. Gregory, et al. (2010). "An LYPSL late domain in the gag protein contributes to the efficient release and replication of Rous sarcoma virus." J Virol **84**(13): 6276-87.
Ding, H., T. J. Green, et al. (2006). "Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus." J Virol **80**(6): 2808-14.
Ding, H., T. J. Green, et al. (2004). "Crystallization and preliminary X-ray analysis of a proteinase-K-resistant domain within the phosphoprotein of vesicular stomatitis virus (Indiana)." Acta Crystallogr D Biol Crystallogr **60**(Pt 11): 2087-90.
Dingwall, C. and R. A. Laskey (1991). "Nuclear targeting sequences--a consensus?" Trends Biochem Sci **16**(12): 478-81.
Doig, A. J. and R. L. Baldwin (1995). "N- and C-capping preferences for all 20 amino acids in alpha-helical peptides." Protein Sci **4**(7): 1325-36.
Doms, R. W., D. S. Keller, et al. (1987). "Role for adenosine triphosphate in regulating the assembly and transport of vesicular stomatitis virus G protein trimers." J Cell Biol **105**(5): 1957-69.
Duan, Y., B. V. Reddy, et al. (2005). "Physicochemical and residue conservation calculations to improve the ranking of protein-protein docking solutions." Protein Sci **14**(2): 316-28.
Dunker, A. K., C. J. Brown, et al. (2002). "Identification and functions of usefully disordered proteins." Adv Protein Chem **62**: 25-49.
Dunker, A. K., Z. Obradovic, et al. (2000). "Intrinsic protein disorder in complete genomes." Genome Inform Ser Workshop Genome Inform **11**: 161-71.

Durrer, P., Y. Gaudin, et al. (1995). "Photolabeling identifies a putative fusion domain in the envelope glycoprotein of rabies and vesicular stomatitis viruses." J Biol Chem **270**(29): 17575-81.
Dyson, H. J. and P. E. Wright (2002). "Coupling of folding and binding for unstructured proteins." Curr Opin Struct Biol **12**(1): 54-60.
Dyson, H. J. and P. E. Wright (2002). "Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance." Adv Protein Chem **62**: 311-40.
Dyson, H. J. and P. E. Wright (2004). "Unfolded proteins and protein folding studied by NMR." Chem Rev **104**(8): 3607-22.
Dyson, H. J. and P. E. Wright (2005). "Intrinsically unstructured proteins and their functions." Nat Rev Mol Cell Biol **6**(3): 197-208.
Eaton, W. A., V. Munoz, et al. (2000). "Fast kinetics and mechanisms in protein folding." Annu Rev Biophys Biomol Struct **29**: 327-59.
Ebert, O., S. Harbaran, et al. (2004). "Systemic therapy of experimental breast cancer metastases by mutant vesicular stomatitis virus in immune-competent mice." Cancer gene therapy **12**(4): 350-358.
Ebert, O., K. Shinozaki, et al. (2003). "Oncolytic vesicular stomatitis virus for treatment of orthotopic hepatocellular carcinoma in immune-competent rats." Cancer research **63**(13): 3605-3605.
Edge, R. E., T. J. Falls, et al. (2008). "A let-7 MicroRNA-sensitive vesicular stomatitis virus demonstrates tumor-specific replication." Molecular Therapy **16**(8): 1437-1443.
Egelman, E. H., S. S. Wu, et al. (1989). "The Sendai virus nucleocapsid exists in at least four different helical states." Journal of Virology **63**(5): 2233-2233.
El-Desouky, A., T. Chi-Wei, et al. (2009). "Cellular and Molecular Interactions of Rhabdoviruses with their Insect and Plant Hosts." Annual Review Of Entomology **54**: 447-468.
Eliezer, D. (2009). "Biophysical characterization of intrinsically disordered proteins." Curr Opin Struct Biol **19**(1): 23-30.
Eliezer, D., J. Chung, et al. (2000). "Native and non-native secondary structure and dynamics in the pH 4 intermediate of apomyoglobin." Biochemistry **39**(11): 2894-901.
Elliott, G. D., R. P. Yeo, et al. (1990). "Strain-variable editing during transcription of the P gene of mumps virus may lead to the generation of non-structural proteins NS1 (V) and NS2." Journal of General Virology **71**(7): 1555-1555.
Emerson, S. U. (1982). "Reconstitution studies detect a single polymerase entry site on the vesicular stomatitis virus genome." Cell **31**(3 Pt 2): 635-42.
Emerson, S. U. and M. Schubert (1987). "Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus." Proc Natl Acad Sci U S A **84**(16): 5655-9.
Emerson, S. U. and R. R. Wagner (1972). "Dissociation and reconstitution of the transcriptase and template activities of vesicular stomatitis B and T virions." J Virol **10**(2): 297-309.
Emerson, S. U. and Y. Yu (1975). "Both NS and L proteins are required for in vitro RNA synthesis by vesicular stomatitis virus." Journal of Virology **15**(6): 1348-1348.
Emerson, S. U. and Y. Yu (1975). "Both NS and L proteins are required for in vitro RNA synthesis by vesicular stomatitis virus." J Virol **15**(6): 1348-56.
Emsley, P., B. Lohkamp, et al. (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.
Eyal, E., R. Najmanovich, et al. (2004). "Importance of solvent accessibility and contact surfaces in modeling side-chain conformations in proteins." J Comput Chem **25**(5): 712-24.
Farrow, N. A., R. Muhandiram, et al. (1994). "Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation." Biochemistry **33**(19): 5984-6003.
Felitsky, D. J., M. A. Lietzow, et al. (2008). "Modeling transient collapsed states of an unfolded protein to provide insights into early folding events." Proc Natl Acad Sci U S A **105**(17): 6278-83.
Ferran, M. C. and J. M. Lucas-Lenard (1997). "The vesicular stomatitis virus matrix protein inhibits transcription from the human beta interferon promoter." Journal of Virology **71**(1): 371-371.
Fields, B. N., D. M. Knipe, et al. (1996). Fields' virology, 3rd edn. New York, Lippincott-Raven Publishers.
Finke, S., K. Brzozka, et al. (2004). "Tracking fluorescence-labeled rabies virus: enhanced green fluorescent protein-tagged phosphoprotein P supports virus gene expression and formation of infectious particles." J Virol **78**(22): 12333-43.
Finke, S. and K. K. Conzelmann (2003). "Dissociation of rabies virus matrix protein functions in regulation of viral RNA synthesis and virus assembly." J Virol **77**(22): 12074-82.
Finke, S. and K. K. Conzelmann (2003). "Dissociation of rabies virus matrix protein functions in regulation of viral RNA synthesis and virus assembly." Journal of Virology **77**(22): 12074-12074.

Finke, S., J. H. Cox, et al. (2000). "Differential transcription attenuation of rabies virus genes by intergenic regions: generation of recombinant viruses overexpressing the polymerase gene." J Virol **74**(16): 7261-9.
Finke, S., R. Mueller-Waldeck, et al. (2003). "Rabies virus matrix protein regulates the balance of virus transcription and replication." Journal of General Virology **84**(6): 1613-1613.
Fitzkee, N. C., P. J. Fleming, et al. (2005). "Are proteins made from a limited parts list?" Trends Biochem Sci **30**(2): 73-80.
Fitzkee, N. C. and G. D. Rose (2004). "Reassessing random-coil statistics in unfolded proteins." Proc Natl Acad Sci U S A **101**(34): 12497-502.
Flanagan, E. B., J. M. Zamparo, et al. (2001). "Rearrangement of the genes of vesicular stomatitis virus eliminates clinical disease in the natural host: new strategy for vaccine development." Journal of Virology **75**(13): 6107-6107.
Flory, P. J. (1953). Principles of polymer chemistry. Ithaca, N.Y., Cornell University Press.
Fooks, A. R. (2005). "Rabies remains a 'neglected disease'." Euro Surveill **10**(11): 211-2.
Fooks, A. R., S. M. Brookes, et al. (2003). "European bat lyssaviruses: an emerging zoonosis." Epidemiol Infect **131**(3): 1029-39.
Förster, F., B. Webb, et al. (2008). "Integration of small-angle X-ray scattering data into structural modeling of proteins and their assemblies." Journal of Molecular Biology **382**(4): 1089-1106.
Franke, D. and D. I. Svergun (2009). "DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering." J. Appl. Cryst. **42**: 342-346.
Fu, Z. F., Y. Zheng, et al. (1994). "Both the N- and the C-terminal domains of the nominal phosphoprotein of rabies virus are involved in binding to the nucleoprotein." Virology **200**(2): 590-7.
Fujinami, R. S., M. B. Oldstone, et al. (1983). "Molecular mimicry in virus infection: crossreaction of measles virus phosphoprotein or of herpes simplex virus protein with human intermediate filaments." Proceedings of the National Academy of Sciences **80**(8): 2346-2346.
Fuxreiter, M., I. Simon, et al. (2004). "Preformed structural elements feature in partner recognition by intrinsically unstructured proteins." J Mol Biol **338**(5): 1015-26.
Galloway, S. E., P. E. Richardson, et al. (2008). "Analysis of a structural homology model of the 2'-O-ribose methyltransferase domain within the vesicular stomatitis virus L protein." Virology **382**(1): 69-82.
Galtier, P. (1881). "Les injections de virus rabique dans le torrent circulatoire ne provoquent pas l'éclosion de la rage et semblent conférer l'immunité. La rage peut être transmise par l'ingestion de matière rabique (Note présentée par M. Bouley)." Comptes rendus de l'Académie des sciences **Tome 93**: p. 284-285.
Gao, Y. and J. Lenard (1995). "Cooperative binding of multimeric phosphoprotein (P) of vesicular stomatitis virus to polymerase (L) and template: pathways of assembly." J Virol **69**(12): 7718-23.
Gao, Y. and J. Lenard (1995). "Multimerization and transcriptional activation of the phosphoprotein (P) of vesicular stomatitis virus by casein kinase-II." Embo J **14**(6): 1240-7.
Garbutt, M., R. Liebscher, et al. (2004). "Properties of replication-competent vesicular stomatitis virus vectors expressing glycoproteins of filoviruses and arenaviruses." Journal of Virology **78**(10): 5458-5458.
Gast, K., D. Zirwer, et al. (2003). "Are there temperature-dependent structural transitions in the "intrinsically unstructured" protein prothymosin alpha?" Eur Biophys J **31**(8): 586-94.
Gast, K., D. Zirwer, et al. (1997). "Ribonuclease T1 has different dimensions in the thermally and chemically denatured states: a dynamic light scattering study." FEBS Lett **403**(3): 245-8.
Gastka, M., J. Horvath, et al. (1996). "Rabies virus binding to the nicotinic acetylcholine receptor alpha subunit demonstrated by virus overlay protein binding assay." J Gen Virol **77 (Pt 10)**: 2437-40.
Gaudier, M., Y. Gaudin, et al. (2002). "Crystal structure of vesicular stomatitis virus matrix protein." EMBO J **21**(12): 2886-92.
Gaudin, Y. (2000). "Rabies virus-induced membrane fusion pathway." The Journal of Cell Biology **150**(3): 601-601.
Gaudin, Y., A. Barge, et al. (1995). "Aggregation of VSV M protein is reversible and mediated by nucleation sites: implications for viral assembly." Virology **206**(1): 28-37.
Gaudin, Y., H. Raux, et al. (1996). "Identification of amino acids controlling the low-pH-induced conformational change of rabies virus glycoprotein." J Virol **70**(11): 7371-8.
Gaudin, Y., R. W. Ruigrok, et al. (1993). "Low-pH conformational changes of rabies virus glycoprotein and their role in membrane fusion." J Virol **67**(3): 1365-72.
Gaudin, Y., R. W. Ruigrok, et al. (1992). "Rabies virus glycoprotein is a trimer." Virology **187**(2): 627-32.
Gaudin, Y., C. Tuffereau, et al. (1991). "Fatty acylation of rabies virus proteins." Virology **184**(1): 441-4.
Gaudin, Y., C. Tuffereau, et al. (1999). "Rabies virus-induced membrane fusion." Mol Membr Biol **16**(1): 21-31.
Gaudin, Y., C. Tuffereau, et al. (1995). "Biological function of the low-pH, fusion-inactive conformation of rabies virus glycoprotein (G): G is transported in a fusion-inactive state-like conformation." J Virol **69**(9): 5528-34.

Gaudin, Y., C. Tuffereau, et al. (1991). "Reversible conformational changes and fusion activity of rabies virus glycoprotein." J Virol **65**(9): 4853-9.
Ge, P., J. Tsao, et al. (2010). "Cryo-EM model of the bullet-shaped vesicular stomatitis virus." Science **327**(5966): 689-93.
Geisbert, T. W., S. Jones, et al. (2005). "Development of a new vaccine for the prevention of Lassa fever." PLoS Medicine **2**(6): 537-537.
Gérard, F., E. Ribeiro, et al. (2007). "Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution." Biochemistry **46**: 10328-10338.
Gérard, F. C. A., E. A. Ribeiro , et al. (2009). "Modular organization of rabies virus phosphoprotein." J. Mol. Biol. **388**: 978-996.
Gigant, B., F. Iseni, et al. (2000). "Neither phosphorylation nor the amino-terminal part of rabies virus phosphoprotein is required for its oligomerization." J Gen Virol **81**(Pt 7): 1757-61.
Gill, D. S., D. Chattopadhyay, et al. (1986). "Identification of a domain within the phosphoprotein of vesicular stomatitis virus that is essential for transcription in vitro." Proc Natl Acad Sci U S A **83**(23): 8873-7.
Glatter, O. and O. Kratky (1982). Small Angle X-ray Scattering. London, UK, Academic Press.
Glodowski, D. R., J. M. Petersen, et al. (2002). "Complex nuclear localization signals in the matrix protein of vesicular stomatitis virus." J Biol Chem **277**(49): 46864-70.
Goddard, T. D. and D. G. Kneller (2003). SPARKY 3. University of California, San Francisco.
Gode, G. R. and N. K. Bhide (1988). "Two rabies deaths after corneal grafts from one donor." Lancet **2**(8614): 791-791.
Golovanov, A. P., G. M. Hautbergue, et al. (2004). "A simple method for improving protein solubility and long-term stability." J Am Chem Soc **126**(29): 8933-9.
Goodin, M., S. Yelton, et al. (2005). "Live-cell imaging of rhabdovirus-induced morphological changes in plant nuclear membranes." Molecular Plant-Microbe Interactions **18**(7): 703-709.
Goodin, M. M., J. Austin, et al. (2001). "Interactions and nuclear import of the N and P proteins of sonchus yellow net virus, a plant nucleorhabdovirus." Journal of Virology **75**(19): 9393-9393.
Graf, W., N. Gerrits, et al. (2002). "Mapping the oculomotor system: the power of transneuronal labelling with rabies virus." European Journal of Neuroscience **15**(9): 1557-1562.
Graham, S. C., R. Assenberg, et al. (2008). "Rhabdovirus matrix protein structures reveal a novel mode of self-association." PLoS Pathog **4**(12): e1000251.
Gray, J. J., S. Moughon, et al. (2003). "Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations." J Mol Biol **331**(1): 281-99.
Grdzelishvili, V. Z., S. Smallwood, et al. (2005). "A single amino acid change in the L-polymerase protein of vesicular stomatitis virus completely abolishes viral mRNA cap methylation." Journal of Virology **79**(12): 7327-7327.
Green, T. J. and M. Luo (2009). "Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P." Proc Natl Acad Sci U S A **106**: 11721-11726.
Green, T. J. and M. Luo (2009). "Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P." Proc Natl Acad Sci U S A **106**(28): 11713-8.
Green, T. J., S. Macpherson, et al. (2000). "Study of the assembly of vesicular stomatitis virus N protein: role of the P protein." J Virol **74**(20): 9515-24.
Green, T. J., X. Zhang, et al. (2006). "Structure of the vesicular stomatitis virus nucleoprotein-RNA complex." Science **313**(5785): 357-60.
Gsponer, J., J. Christodoulou, et al. (2008). "A coupled equilibrium shift mechanism in calmodulin-mediated signal transduction." Structure **16**(5): 736-46.
Gupta, A. K. and A. K. Banerjee (1997). "Expression and purification of vesicular stomatitis virus N-P complex from Escherichia coli: role in genome RNA transcription and replication in vitro." J Virol **71**(6): 4264-71.
Gupta, A. K. and A. K. Banerjee (1997). "Expression and purification of vesicular stomatitis virus NP complex from Escherichia coli: role in genome RNA transcription and replication in vitro." Journal of Virology **71**(6): 4264-4264.
Gupta, A. K., D. Blondel, et al. (2000). "The phosphoprotein of rabies virus is phosphorylated by a unique cellular protein kinase and specific isomers of protein kinase C." J Virol **74**(1): 91-8.
Gupta, A. K., M. Mathur, et al. (2002). "Unique capping activity of the recombinant RNA polymerase (L) of vesicular stomatitis virus: association of cellular capping enzyme with the L protein* 1." Biochemical and biophysical research communications **293**(1): 264-268.
Gupta, A. K., D. Shaji, et al. (2003). "Identification of a novel tripartite complex involved in replication of vesicular stomatitis virus genome RNA." J Virol **77**(1): 732-8.

Habchi, J., L. Mamelli, et al. (2010). "Structural disorder within Henipavirus nucleoprotein and phosphoprotein: from predictions to experimental assessment." PLoS One **5**(7): e11684.
Haglund, K., I. Leiner, et al. (2002). "High-level primary CD8+ T-cell response to human immunodeficiency virus type 1 Gag and Env generated by vaccination with recombinant vesicular stomatitis viruses." Journal of Virology **76**(6): 2730-2730.
Halpin, K., B. Bankamp, et al. (2004). "Nipah virus conforms to the rule of six in a minigenome replication assay." Journal of General Virology **85**(3): 701-701.
Hanlon, C. A., I. V. Kuzmin, et al. (2005). "Efficacy of rabies biologics against new lyssaviruses from Eurasia." Virus Res **111**(1): 44-54.
Hanson, R. P. (1952). "The natural history of vesicular stomatitis." Bacteriol Rev **16**(3): 179-204.
Harrison, R. W. (1993). "Stiffness and Energy Conservation in Molecular Dynamics: an Improved Integrator." J. Comp. Chem. **14** 1112-1122.
Harty, R. N., M. E. Brown, et al. (2001). "Rhabdoviruses and the cellular ubiquitin-proteasome system: a budding interaction." Journal of Virology **75**(22): 10623-10623.
Harty, R. N., J. Paragas, et al. (1999). "A proline-rich motif within the matrix protein of vesicular stomatitis virus and rabies virus interacts with WW domains of cellular proteins: implications for viral budding." J Virol **73**(4): 2921-9.
Harty, R. N., J. Paragas, et al. (1999). "A proline-rich motif within the matrix protein of vesicular stomatitis virus and rabies virus interacts with WW domains of cellular proteins: implications for viral budding." Journal of Virology **73**(4): 2921-2921.
Haupt, W. (1999). "Rabies-risk of exposure and current trends in prevention of human cases." Vaccine **17**(13-14): 1742-1749.
He, B., K. Wang, et al. (2009). "Predicting intrinsic disorder in proteins: an overview." Cell Res **19**(8): 929-49.
He, X., J. Zhou, et al. (2008). "Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus." Nature **454**(7208): 1123-6.
Heggeness, M. H., A. Scheid, et al. (1980). "Conformation of the helical nucleocapsids of paramyxoviruses and vesicular stomatitis virus: reversible coiling and uncoiling induced by changes in salt concentration." Proc Natl Acad Sci U S A **77**(5): 2631-5.
Heldwein, E. E., H. Lou, et al. (2006). "Crystal structure of glycoprotein B from herpes simplex virus 1." Science **313**(5784): 217-20.
Heller, W. T. (2004). "Influence of multiple well defined conformations on small-angle scattering of proteins in solution." Acta Crystallographica Section D: Biological Crystallography **61**(1): 33-44.
Hemachudha, T., J. Laothamatas, et al. (2002). "Human rabies: a disease of complex neuropathogenetic mechanisms and diagnostic challenges." Lancet Neurol **1**(2): 101-9.
Hemachudha, T., J. Laothamatas, et al. (2002). "Human rabies: a disease of complex neuropathogenetic mechanisms and diagnostic challenges." The Lancet Neurology **1**(2): 101-109.
Hemachudha, T., B. Sunsaneewitayakul, et al. (2006). "Failure of therapeutic coma and ketamine for therapy of human rabies." Journal of neurovirology **12**(5): 407-409.
Her, L. S., E. Lund, et al. (1997). "Inhibition of Ran guanosine triphosphatase-dependent nuclear transport by the matrix protein of vesicular stomatitis virus." Science **276**(5320): 1845-8.
Hess, B., H. Bekker, et al. (1997). "LINCS: a linear constraint solver for molecular simulations." Journal of Computational Chemistry **18**(12): 1463-1472.
Hess, B., C. Kutzner, et al. (2008). "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation." J. Chem. Theory Comput. **4**: 435-447.
Hill, V. M., L. Marnell, et al. (1981). "In vitro replication and assembly of vesicular stomatitis virus nucleocapsids." Virology **113**(1): 109-18.
Hodak, H., A. Wohlkönig, et al. (2008). "The peptidyl-prolyl isomerase and chaperone Par27 of *Bordetella pertussis* as the prototype for a new group of parvulins." J Mol Biol **376**(2): 414-26.
Hogenhout, S. A., M. G. Redinbaugh, et al. (2003). "Plant and animal rhabdovirus host range: a bug's view." Trends Microbiol **11**(6): 264-71.
Holmes, D. E. and S. A. Moyer (2002). "The phosphoprotein (P) binding site resides in the N terminus of the L polymerase subunit of sendai virus." J Virol **76**(6): 3078-83.
Horikami, S. M. and S. A. Moyer (1982). "Host range mutants of vesicular stomatitis virus defective in in vitro RNA methylation." Proceedings of the National Academy of Sciences of the United States of America **79**(24): 7694-7694.
Hornung, V., J. Ellegast, et al. (2006). "5'-Triphosphate RNA is the ligand for RIG-I." Science **314**(5801): 994-7.
Howard, M. and G. Wertz (1989). "Vesicular stomatitis virus RNA replication: a role for the NS protein." J Gen Virol **70 (Pt 10)**: 2683-94.
Howatson, A. F. and G. F. Whitmore (1962). "The development and structure of vesicular stomatitis virus." Virology **16**: 466-78.

Huang, A. S., D. Baltimore, et al. (1970). "Ribonucleic acid synthesis of vesicular stomatitis virus* 1:: III. Multiple complementary messenger RNA molecules." Virology **42**(4): 946-957.
Hudson, L. D., C. Condra, et al. (1986). "Cloning and expression of a viral phosphoprotein: structure suggests vesicular stomatitis virus NS may function by mimicking an RNA template." J Gen Virol **67 (Pt 8)**: 1571-9.
Hughson, F. M. (1997). "Enveloped viruses: A common mode of membrane fusion?" Current Biology **7**(9): R565-R569-R565-R569.
Hunt, L. A. and D. F. Summers (1976). "Association of vesicular stomatitis virus proteins with HeLa cell membranes and released virus." Journal of Virology **20**(3): 637-637.
Huntley, C. C., B. P. De, et al. (1997). "Phosphorylation of Sendai virus phosphoprotein by cellular protein kinase C zeta." J Biol Chem **272**(26): 16578-84.
Hwang, L. N., N. Englund, et al. (1999). "Optimal replication activity of vesicular stomatitis virus RNA polymerase requires phosphorylation of a residue(s) at carboxy-terminal domain II of its accessory subunit, phosphoprotein P." J Virol **73**(7): 5613-20.
Irie, T., E. Carnero, et al. (2007). "Modifications of the PSAP region of the matrix protein lead to attenuation of vesicular stomatitis virus in vitro and in vivo." J Gen Virol **88**(Pt 9): 2559-67.
Irie, T. and R. N. Harty (2005). "L-domain flanking sequences are important for host interactions and efficient budding of vesicular stomatitis virus recombinants." Journal of Virology **79**(20): 12617-12617.
Irie, T., J. M. Licata, et al. (2004). "Functional analysis of late-budding domain activity associated with the PSAP motif within the vesicular stomatitis virus M protein." J Virol **78**(14): 7823-7.
Is, R. (2005). "Recovery of a patient from clinical rabies--Wisconsin, 2004." JAMA **293**(6): 669-669.
Iseni, F., A. Barge, et al. (1998). "Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures." J Gen Virol **79 (Pt 12)**: 2909-19.
Iseni, F., F. Baudin, et al. (2000). "Structure of the RNA inside the vesicular stomatitis virus nucleocapsid." Rna **6**(2): 270-81.
Iseni, F., F. Baudin, et al. (2002). "Chemical modification of nucleotide bases and mRNA editing depend on hexamer or nucleoprotein phase in Sendai virus nucleocapsids." Rna **8**(8): 1056-67.
Ito, N., G. W. Moseley, et al. (2010). "Role of interferon antagonist activity of rabies virus phosphoprotein in viral pathogenicity." J Virol **84**(13): 6699-710.
Ivanov, I., T. Crepin, et al. (2010). "Structure of the dimerisation domain of the rabies virus phosphoprotein." J. Virol. **84**: 3707-3710.
Iverson, L. E. and J. K. Rose (1981). "Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription." Cell **23**(2): 477-84.
Iverson, L. E. and J. K. Rose (1982). "Sequential synthesis of 5'-proximal vesicular stomatitis virus mRNA sequences." J Virol **44**(1): 356-65.
Iyer, A. V., B. Pahar, et al. (2009). "Recombinant vesicular stomatitis virus-based west Nile vaccine elicits strong humoral and cellular immune responses and protects mice against lethal challenge with the virulent west Nile virus strain LSU-AR01." Vaccine **27**(6): 893-903.
Jackson, A. C. (2009). "Update on rabies diagnosis and treatment." Current infectious disease reports **11**(4): 296-301.
Jacob, Y., H. Badrane, et al. (2000). "Cytoplasmic dynein LC8 interacts with lyssavirus phosphoprotein." J Virol **74**(21): 10217-22.
Jacob, Y., E. Real, et al. (2001). "Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains." J Virol **75**(20): 9613-22.
Jarvet, J., P. Damberg, et al. (2003). "A left-handed 31 helical conformation in the Alzheimer A [beta](12-28) peptide." FEBS letters **555**(2): 371-374.
Jayakar, H. R., K. G. Murti, et al. (2000). "Mutations in the PPPY motif of vesicular stomatitis virus matrix protein reduce virus budding by inhibiting a late step in virion release." Journal of Virology **74**(21): 9818-9818.
Jeganathan, S., M. von Bergen, et al. (2008). "The natively unfolded character of tau and its aggregation to Alzheimer-like paired helical filaments." Biochemistry **47**(40): 10526-39.
Jensen, M. R., K. Houben, et al. (2008). "Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein." J Am Chem Soc **130**(25): 8055-61.
Jensen, M. R., P. R. Markwick, et al. (2009). "Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings." Structure **17**(9): 1169-85.
Jensen, M. R., L. Salmon, et al. (2010). "Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts." J Am Chem Soc **132**(4): 1270-2.

Jha, A. K., A. Colubri, et al. (2005). "Statistical coil model of the unfolded state: resolving the reconciliation problem." Proc Natl Acad Sci U S A **102**(37): 13099-104.
Jha, A. K., A. Colubri, et al. (2005). "Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library." Biochemistry **44**(28): 9691-702.
Johannsdottir, H. K., R. Mancini, et al. (2009). "Host cell factors and functions involved in vesicular stomatitis virus entry." Journal of Virology **83**(1): 440-440.
Johnson, K. M., R. B. Tesh, et al. (1969). "Epidemiology of vesicular stomatitis virus: some new data and a hypothesis for transmission of the Indian serotype." J Am Vet Med Assoc **155**(12): 2133-40.
Johnson, K. M., J. E. Vogel, et al. (1966). "Clinical and serological response to laboratory-acquired human infection by Indiana type vesicular stomatitis virus (VSV)." The American Journal of Tropical Medicine and Hygiene **15**(2): 244-244.
Jones, O. W. and P. Berg (1966). "Studies on the binding of RNA polymerase to polynucleotides* 1." Journal of Molecular Biology **22**(2): 199-209.
Jones, S. and J. M. Thornton (1996). "Principles of protein-protein interactions." Proc Natl Acad Sci U S A **93**(1): 13-20.
Jung, Y. S. and M. Zweckstetter (2004). "Mars – robust automatic backbone assignment of proteins." J. Biomolecular NMR **30**: 11-23.
Kalmakoff, J., L. J. Lewandowski, et al. (1969). "Comparison of the ribonucleic Acid subunits of reovirus, cytoplasmic polyhedrosis virus, and wound tumor virus." J Virol **4**(6): 851-6.
Kalvodova, L., J. L. Sampaio, et al. (2009). "The Lipidomes of Vesicular Stomatitis Virus, Semliki Forest Virus, and the Host Plasma Membrane Analyzed by Quantitative Shotgun Mass Spectrometry." Journal of Virology **83**(16): 7996-7996.
Kaminski, G. A., R. A. Friesner, et al. (2001). "Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides." J. Phys. Chem. B **105**: 6474-6487

Kaptur, P. E., M. O. McKenzie, et al. (1995). "Assembly functions of vesicular stomatitis virus matrix protein are not disrupted by mutations at major sites of phosphorylation." Virology **206**(2): 894-903.
Karlin, D., F. Ferron, et al. (2003). "Structural disorder and modular organization in Paramyxovirinae N and P." J Gen Virol **84**(Pt 12): 3239-52.
Katze, M. G., J. L. Fornek, et al. (2008). "Innate immune modulation by RNA viruses: emerging insights from functional genomics." Nat Rev Immunol **8**(8): 644-54.
Kawai, A., H. Toriumi, et al. (1999). "Nucleocapsid formation and/or subsequent conformational change of rabies virus nucleoprotein (N) is a prerequisite step for acquiring the phosphatase-sensitive epitope of monoclonal antibody 5-2-26." Virology **263**(2): 395-407.
Keene, J. D., B. J. Thornton, et al. (1981). "Sequence-specific contacts between the RNA polymerase of vesicular stomatitis virus and the leader RNA gene." Proc Natl Acad Sci U S A **78**(10): 6191-5.
Kelly, R. M. and P. L. Strick (2000). "Rabies as a transneuronal tracer of circuits in the central nervous system." Journal of neuroscience methods **103**(1): 63-71.
Kelly, R. M. and P. L. Strick (2004). "Macro-architecture of basal ganglia loops with the cerebral cortex: use of rabies virus to reveal multisynaptic circuits." Progress in Brain Research **143**: 447-459.
Kiley, M. P. and R. R. Wagner (1972). "Ribonucleic acid species of intracellular nucleocapsids and released virions of vesicular stomatitis virus." Journal of Virology **10**(2): 244-244.
Kjaergaard, M., A. B. Norholm, et al. (2010). "Temperature-dependent structural changes in intrinsically disordered proteins: formation of alpha-helices or loss of polyproline II?" Protein Sci **19**(8): 1555-64.
Klingen, Y., K. K. Conzelmann, et al. (2008). "Double-labeled rabies virus: live tracking of enveloped virus transport." J Virol **82**(1): 237-45.
Knipe, D. M. (2006). "Fields virology (Two-Volume Set, with CD-ROM)."
Knobel, D. L., S. Cleaveland, et al. (2005). "Re-evaluating the burden of rabies in Africa and Asia." Bull World Health Organ **83**(5): 360-8.
Kobayashi, Y., G. Sato, et al. (2008). "Molecular and geographic analyses of vampire bat-transmitted cattle rabies in central Brazil." BMC Vet Res **4**: 44.
Koch, M. H., P. Vachette, et al. (2003). "Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution." Q Rev Biophys **36**(2): 147-227.
Kohn, J. E., I. S. Millett, et al. (2004). "Random-coil behavior and the dimensions of chemically unfolded proteins." Proc Natl Acad Sci U S A **101**(34): 12491-6.
Kolakofsky, D., P. Le Mercier, et al. (2004). "Viral RNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis." Virology **318**(2): 463-73.
Kolakofsky, D., T. Pelet, et al. (1998). "Paramyxovirus RNA synthesis and the requirement for hexamer genome length: the rule of six revisited." Journal of Virology **72**(2): 891-891.

Kolakofsky, D., L. Roux, et al. (2005). "Paramyxovirus mRNA editing, the'rule of six'and error catastrophe: a hypothesis." Journal of General Virology **86**(7): 1869-1869.
Komarova, A. V., E. Real, et al. (2007). "Rabies virus matrix protein interplay with eIF3, new insights into rabies virus pathogenesis." Nucleic Acids Res **35**(5): 1522-32.
Komarova, A. V., E. Real, et al. (2007). "Rabies virus matrix protein interplay with eIF3, new insights into rabies virus pathogenesis." Nucleic acids research.
Konarev, P., M. Petoukhov, et al. (2006). "ATSAS 2.1, a program package for small-angle scattering data analysis." J. Appl. Cryst. **39**: 277-286.
Konarev, P. V., V. V. Volkov, et al. (2003). "PRIMUS: a Windows PC-based system for small-angle scattering data analysis." J. Appl. Cryst. **36**: 1277-1282.
Kortemme, T., D. E. Kim, et al. (2004). "Computational alanine scanning of protein-protein interfaces." Science's STKE **2004**(219).
Kozak, M. (2005). "Glucose isomerase from Streptomyces rubiginosus – potential molecular weight standard for small-angle X-ray scattering." J. Appl. Cryst. **38**: 555-558.
Kretzschmar, E., L. Buonocore, et al. (1997). "High-efficiency incorporation of functional influenza virus glycoproteins into recombinant vesicular stomatitis viruses." Journal of Virology **71**(8): 5982-5982.
Krippahl, L., J. J. Moura, et al. (2003). "Modeling protein complexes with BiGGER." Proteins **52**(1): 19-23.
Krissinel, E. and K. Henrick (2004). "Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2256-68.
Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." J Mol Biol **372**(3): 774-97.
Krol, M., R. A. Chaleil, et al. (2007). "Implicit flexibility in protein docking: cross-docking and local refinement." Proteins **69**(4): 750-7.
Krol, M., A. L. Tournier, et al. (2007). "Flexible relaxation of rigid-body docking solutions." Proteins **68**(1): 159-69.
Kurath, G., K. G. Ahern, et al. (1985). "Molecular cloning of the six mRNA species of infectious hematopoietic necrosis virus, a fish rhabdovirus, and gene order determination by R-loop mapping." J Virol **53**(2): 469-76.
Kurath, G. and J. C. Leong (1985). "Characterization of infectious hematopoietic necrosis virus mRNA species reveals a nonvirion rhabdovirus protein." J Virol **53**(2): 462-8.
Kurilla, M. G., H. Piwnica-Worms, et al. (1982). "Rapid and transient localization of the leader RNA of vesicular stomatitis virus in the nuclei of infected cells." Proc Natl Acad Sci U S A **79**(17): 5240-4.
Kurotani, A., K. Kiyotani, et al. (1998). "Sendai virus C proteins are categorically nonessential gene products but silencing their expression severely impairs viral replication and pathogenesis." Genes to cells **3**(2): 111-124.
Kuzmin, I. V., R. Franka, et al. (2008). "Experimental infection of big brown bats (Eptesicus fuscus) with West Caucasian bat virus (WCBV)." Dev Biol (Basel) **131**: 327-37.
Kuzmin, I. V., A. E. Mayer, et al. (2010). "Shimoni bat virus, a new representative of the Lyssavirus genus." Virus Res **149**(2): 197-210.
Kuzmin, I. V., M. Niezgoda, et al. (2008). "Possible emergence of West Caucasian bat virus in Africa." Emerg Infect Dis **14**(12): 1887-9.
Kuzmin, I. V., X. Wu, et al. (2008). "Complete genomes of Aravan, Khujand, Irkut and West Caucasian bat viruses, with special attention to the polymerase gene and non-coding regions." Virus Res **136**(1-2): 81-90.
La Scola, B., S. Audic, et al. (2003). "A giant virus in amoebae." Science **299**(5615): 2033.
Lacroix, E., A. R. Viguera, et al. (1998). "Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters." J Mol Biol **284**(1): 173-91.
Lahaye, X., A. Vidy, et al. (2009). "Functional characterization of Negri bodies (NBs) in rabies virus-infected cells: Evidence that NBs are sites of viral transcription and replication." J Virol **83**(16): 7948-58.
Lange, O. F., N. A. Lakomek, et al. (2008). "Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution." Science **320**(5882): 1471-5.
Laothamatas, J., S. Wacharapluesadee, et al. (2008). "Furious and paralytic rabies of canine origin: Neuroimaging with virological and cytokine studies." Journal of neurovirology **14**(2): 119-129.
Laskowski, R. A., M. W. Macarthur, et al. (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." J. Appl. Cryst. **26**: 283-291.
Lentz, T. L., E. Hawrot, et al. (1987). "Synthetic peptides corresponding to sequences of snake venom neurotoxins and rabies virus glycoprotein bind to the nicotinic acetylcholine receptor." Proteins **2**(4): 298-307.

Leopold, P. L. and K. K. Pfister (2006). "Viral strategies for intracellular trafficking: motors and microtubules." Traffic **7**(5): 516-23.
Lesburg, C. A., M. B. Cable, et al. (1999). "Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site." Nat Struct Biol **6**(10): 937-43.
Lescop, E., P. Schanda, et al. (2007). "A set of BEST triple-resonance experiments for time-optimized protein resonance assignment." J Magn Reson **187**(1): 163-9.
Leslie, M. J., S. Messenger, et al. (2006). "Bat-associated rabies virus in Skunks." Emerg Infect Dis **12**(8): 1274-7.
Lewandowski, L. J., J. Kalmakoff, et al. (1969). "Characterization of a Ribonucleic Acid Polymerase Activity Associated with Purified Cytoplasmic Polyhedrosis Virus of the Silkworm Bombyx mori." J Virol **4**(6): 857-65.
Leyrat, C., F. C. Gerard, et al. (2010). "Structural disorder in proteins of the rhabdoviridae replication complex." Protein Pept Lett **17**(8): 979-87.
Li, J., J. S. Chorba, et al. (2007). "Vesicular stomatitis viruses resistant to the methylase inhibitor sinefungin upregulate RNA synthesis and reveal mutations that affect mRNA cap methylation." Journal of Virology **81**(8): 4104-4104.
Li, J., E. C. Fontaine-Rodriguez, et al. (2005). "Amino acid residues within conserved domain VI of the vesicular stomatitis virus large polymerase protein essential for mRNA cap methyltransferase activity." Journal of Virology **79**(21): 13373-13373.
Li, J., A. Rahmeh, et al. (2009). "Opposing effects of inhibiting cap addition and cap methylation on polyadenylation during vesicular stomatitis virus mRNA synthesis." Journal of Virology **83**(4): 1930-1930.
Li, J., J. T. Wang, et al. (2006). "A unique strategy for mRNA cap methylation used by vesicular stomatitis virus." Proceedings of the National Academy of Sciences **103**(22): 8493-8493.
Li, T. and A. K. Pattnaik (1999). "Overlapping signals for transcription and replication at the 3' terminus of the vesicular stomatitis virus genome." J Virol **73**(1): 444-52.
Li, T. and A. K. Pattnaik (1999). "Overlapping signals for transcription and replication at the 3'terminus of the vesicular stomatitis virus genome." Journal of Virology **73**(1): 444-444.
Liang, S., S. Liu, et al. (2007). "A simple reference state makes a significant improvement in near-native selections from structurally refined docking decoys." Proteins **69**(2): 244-53.
Liang, S., G. Wang, et al. (2008). "Refining near-native protein-protein docking decoys by local resampling and energy minimization." Proteins.
Liao, J. B., J. Publicover, et al. (2008). "Single-dose, therapeutic vaccination of mice with vesicular stomatitis virus expressing human papillomavirus type 16 E7 protein." Clinical and Vaccine Immunology **15**(5): 817-817.
Lichty, B. D., A. T. Power, et al. (2004). "Vesicular stomatitis virus: re-inventing the bullet." Trends Mol Med **10**(5): 210-6.
Lichty, B. D., D. F. Stojdl, et al. (2004). "Vesicular stomatitis virus: a potential therapeutic virus for the treatment of hematologic malignancy." Hum Gene Ther **15**(9): 821-31.
Liu, Y., L. Cocka, et al. (2010). "Conserved motifs within Ebola and Marburg virus VP40 proteins are important for stability, localization, and subsequent budding of virus-like particles." J Virol **84**(5): 2294-303.
Llorente, M. T., B. Garcia-Barreno, et al. (2006). "Structural analysis of the human respiratory syncytial virus phosphoprotein: characterization of an alpha-helical domain involved in oligomerization." J Gen Virol **87**(Pt 1): 159-69.
Lo Conte, L., C. Chothia, et al. (1999). "The atomic structure of protein-protein recognition sites." J Mol Biol **285**(5): 2177-98.
Lotti, L. V., M. R. Torrisi, et al. (1992). "Immunocytochemical analysis of the transfer of vesicular stomatitis virus G glycoprotein from the intermediate compartment to the Golgi complex." Journal of Cell Biology **118**(1): 43-43.
Luan, P., L. Yang, et al. (1995). "Formation of membrane domains created during the budding of vesicular stomatitis virus. A model for selective lipid and protein sorting in biological membranes." Biochemistry **34**(31): 9874-83.
Luo, M., T. J. Green, et al. (2007). "Conserved characteristics of the rhabdovirus nucleoprotein." Virus Research **129**(2): 246-251.
Luo, M., T. J. Green, et al. (2007). "Conserved characteristics of the rhabdovirus nucleoprotein." Virus Res **129**(2): 246-51.
Luo, P. and R. L. Baldwin (1997). "Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water." Biochemistry **36**: 8413-8421.
Lwoff, A. (1957). "The concept of virus." Microbiology **17**(2): 239-239.

Ma, J. and M. Ptashne (1987). "Deletion analysis of GAL4 defines two transcriptional activating segments." Cell **48**(5): 847-53.
Madhusudana, S. N., D. Nagaraj, et al. (2002). "Partial recovery from rabies in a six-year-old girl." International journal of infectious diseases: IJID: official publication of the International Society for Infectious Diseases **6**(1): 85-85.
Malm, J., M. Jonsson, et al. (2007). "Structural properties of semenogelin I." FEBS J **274**(17): 4503-10.
Mao, A. H., S. L. Crick, et al. (2010). "Net charge per residue modulates conformational ensembles of intrinsically disordered proteins." Proc Natl Acad Sci U S A **107**(18): 8183-8.
Marriott, A. C. (2005). "Complete genome sequences of Chandipura and Isfahan vesiculoviruses." Archives of virology **150**(4): 671-680.
Marsh, J. A., V. K. Singh, et al. (2006). "Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation." Protein Sci **15**(12): 2795-804.
Martins, C. R. F., J. A. Johnson, et al. (1998). "Sonchus yellow net rhabdovirus nuclear viroplasms contain polymerase-associated proteins." Journal of virology **72**(7): 5669-5669.
Masters, P. S. and A. K. Banerjee (1986). "Phosphoprotein NS of vesicular stomatitis virus: phosphorylated states and transcriptional activities of intracellular and virion forms." Virology **154**(2): 259-70.
Masters, P. S. and A. K. Banerjee (1988). "Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA." J Virol **62**(8): 2658-64.
Mathur, M., T. Das, et al. (1997). "Display of disparate transcription phenotype by the phosphorylation negative P protein mutants of vesicular stomatitis virus, Indiana serotype, expressed in E. coli and eucaryotic cells." Gene Expr **6**(5): 275-86.
Matlin, K. S., H. Reggio, et al. (1982). "Pathway of vesicular stomatitis virus entry leading to infection." J Mol Biol **156**(3): 609-31.
Mavrakis, M., F. Iseni, et al. (2003). "Isolation and characterisation of the rabies virus N degrees-P complex produced in insect cells." Virology **305**(2): 406-14.
Mavrakis, M., F. d. r. Iseni, et al. (2003). "Isolation and Characterisation of the Rabies Virus N°-P Complex Produced in Insect Cells." Virology **305**(2): 406-414.
Mavrakis, M., L. Kolesnikova, et al. (2002). "Morphology of Marburg virus NP-RNA." Virology **296**(2): 300-7.
Mavrakis, M., A. A. McCarthy, et al. (2004). "Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus." J Mol Biol **343**(4): 819-31.
Mavrakis, M., S. Mehouas, et al. (2006). "Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA." Virology **349**(2): 422-9.
McColl, K. A., N. Tordo, et al. (2000). "Bat lyssavirus infections." Rev Sci Tech **19**(1): 177-96.
McCoy, A. J., R. W. Grosse-Kunstleve, et al. (2007). "Phaser crystallographic software." J Appl Crystallogr **40**(Pt 4): 658-674.
McCreedy, B. J., Jr., K. P. McKinnon, et al. (1990). "Solubility of vesicular stomatitis virus M protein in the cytosol of infected cells or isolated from virions." J Virol **64**(2): 902-6.
McSharry, J. J. and R. R. Wagner (1971). "Carbohydrate composition of vesicular stomatitis virus." J Virol **7**(3): 412-5.
Mebatsion, T. (2001). "Extensive attenuation of rabies virus by simultaneously modifying the dynein light chain binding site in the P protein and replacing Arg333 in the G protein." J Virol **75**(23): 11496-502.
Mebatsion, T., M. Konig, et al. (1996). "Budding of rabies virus particles in the absence of the spike glycoprotein." Cell **84**(6): 941-51.
Mebatsion, T., F. Weiland, et al. (1999). "Matrix protein of rabies virus is responsible for the assembly and budding of bullet-shaped particles and interacts with the transmembrane spike glycoprotein G." J Virol **73**(1): 242-50.
Mebatsion, T., F. Weiland, et al. (1999). "Matrix protein of rabies virus is responsible for the assembly and budding of bullet-shaped particles and interacts with the transmembrane spike glycoprotein G." Journal of Virology **73**(1): 242-242.
Mello, C. C. and D. Barrick (2003). "Measuring the stability of partly folded proteins using TMAO." Protein Sci **12**(7): 1522-9.
Mellon, M. G. and S. U. Emerson (1978). "Rebinding of transcriptase components (L and NS proteins) to the nucleocapsid template of vesicular stomatitis virus." J Virol **27**(3): 560-7.
Meng, E. C., E. F. Pettersen, et al. (2006). "Tools for integrated sequence-structure analysis with UCSF Chimera." BMC Bioinformatics **7**: 339.
Mertens, H. D. and D. I. Svergun (2010). "Structural characterization of proteins and complexes using small-angle X-ray solution scattering." J Struct Biol **172**(1): 128-41.
Mészáros, B., M. Tompa, et al. (2007). "Molecular Principles of the Interactions of Disordered Proteins." J. Mol. Biol. **372**: 549-561.

Mezei, M., P. J. Fleming, et al. (2004). "Polyproline II helix is the preferred conformation for unfolded polyalanine in water." Proteins **55**(3): 502-7.
Mihel, J., M. Sikic, et al. (2008). "PSAIA - protein structure and interaction analyzer." BMC Struct Biol **8**: 21.
Miller, W. G. and C. V. Goebel (1968). "Dimensions of protein random coils." Biochemistry **7**(11): 3925-35.
Millett, I. S., S. Doniach, et al. (2002). "Toward a taxonomy of the denatured state: small angle scattering studies of unfolded proteins." Adv Protein Chem **62**: 241-62.
Mire, C. E., D. Dube, et al. (2009). "Glycoprotein-Dependent Acidification of Vesicular Stomatitis Virus Enhances Release of Matrix Protein." Journal of Virology **83**(23): 12139-12139.
Mire, C. E., J. M. White, et al. (2010). "A Spatio-Temporal Analysis of Matrix Protein and Nucleocapsid Trafficking during Vesicular Stomatitis Virus Uncoating." PLoS pathogens **6**(7): e1000994-e1000994.
Mittag, T., S. Orlicky, et al. (2008). "Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor." Proc Natl Acad Sci U S A **105**(46): 17772-7.
Miyamoto, K. and S. Matsumoto (1965). "The nature of the Negri body." J Cell Biol **27**(3): 677-82.
Miyamoto, S. and P. A. Kollman (1992). "SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models." Journal of Computational Chemistry **13**(8): 952-962.
Moglich, A., K. Joder, et al. (2006). "End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation." Proc Natl Acad Sci U S A **103**(33): 12394-9.
Mohan, A., C. J. Oldfield, et al. (2006). "Analysis of molecular recognition features (MoRFs)." J Mol Biol **362**(5): 1043-59.
Morcuende, S., J. M. Delgado-Garcia, et al. (2002). "Neuronal premotor networks involved in eyelid responses: retrograde transneuronal tracing with rabies virus from the orbicularis oculi muscle in the rat." Journal of Neuroscience **22**(20): 8808-8808.
Moschovakis, A. K., G. G. Gregoriou, et al. (2004). "Oculomotor areas of the primate frontal lobes: a transneuronal transfer of rabies virus and [14C]-2-deoxyglucose functional imaging study." Journal of Neuroscience **24**(25): 5726-5726.
Moseley, G. W., R. P. Filmer, et al. (2007). "Nucleocytoplasmic distribution of rabies virus P-protein is regulated by phosphorylation adjacent to C-terminal nuclear import and export signals." Biochemistry **46**(43): 12053-61.
Moseley, G. W., X. Lahaye, et al. (2009). "Dual modes of rabies P-protein association with microtubules: a novel strategy to suppress the antiviral response." J Cell Sci **122**(Pt 20): 3652-62.
Moseley, G. W., D. M. Roth, et al. (2007). "Dynein light chain association sequences can facilitate nuclear protein import." Mol Biol Cell **18**(8): 3204-13.
Moyer, S. A., G. Abraham, et al. (1975). "Methylated and blocked 5' termini in vesicular stomatitis virus in vivo mRNAs." Cell **5**(1): 59-67.
Moyer, S. A., S. C. Baker, et al. (1986). "Tubulin: a factor necessary for the synthesis of both Sendai virus and vesicular stomatitis virus RNAs." Proc Natl Acad Sci U S A **83**(15): 5405-9.
Moyer, S. A., S. Smallwood-Kentro, et al. (1991). "Assembly and transcription of synthetic vesicular stomatitis virus nucleocapsids." J Virol **65**(5): 2170-8.
Mrak, R. E. and L. Young (1994). "Rabies encephalitis in humans: pathology, pathogenesis and pathophysiology." Journal of Neuropathology & Experimental Neurology **53**(1): 1-1.
Muller-Späth, S., A. Soranno, et al. (2010). "Charge interactions can dominate the dimensions of intrinsically disordered proteins." Proc Natl Acad Sci U S A **107**: 14609-14614.
Munoz, V. and L. Serrano (1994). "Elucidating the folding problem of helical peptides using empirical parameters." Nat Struct Biol **1**(6): 399-409.
Munyon, W., E. Paoletti, et al. (1967). "RNA polymerase activity in purified infectious vaccinia virus." Proceedings of the National Academy of Sciences of the United States of America: 2280-2287.
Murshudov, G. N., A. A. Vagin, et al. (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallographica Section D: Biological Crystallography **53**(3): 240-255.
Negri, A. (1903). "Beitrag zum Studium der Aetiologie der Tollwuth." Medical Microbiology and Immunology **43**(1): 507-528.
Nettels, D., S. Muller-Spath, et al. (2009). "Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins." Proc Natl Acad Sci U S A **106**(49): 20740-5.
Neumann, P., D. Lieber, et al. (2009). "Crystal structure of the Borna disease virus matrix protein (BDV-M) reveals ssRNA binding properties." Proc Natl Acad Sci U S A **106**(10): 3710-5.
Newcomb, W. W. and J. C. Brown (1981). "Role of the vesicular stomatitis virus matrix protein in maintaining the viral nucleocapsid in the condensed form found in native virions." Journal of Virology **39**(1): 295-295.
Noppert, A., K. Gast, et al. (1996). "Reduced-denatured ribonuclease A is not in a compact state." FEBS Lett **380**(1-2): 179-82.

Ogino, T. and A. K. Banerjee (2007). "Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus." Mol Cell **25**(1): 85-97.
Ogino, T., S. P. Yadav, et al. (2010). "Histidine-mediated RNA transfer to GDP for unique mRNA capping by vesicular stomatitis virus RNA polymerase." Proceedings of the National Academy of Sciences **107**(8): 3463-3463.
Ohara, S., K. Inoue, et al. (2009). "Dual transneuronal tracing in the rat entorhinal-hippocampal circuit by intracerebral injection of recombinant rabies virus vectors." Frontiers in Neuroanatomy **3**.
Oldfield, C. J., Y. Cheng, et al. (2005). "Comparing and combining predictors of mostly disordered proteins." Biochemistry **44**(6): 1989-2000.
Oldfield, C. J., Y. Cheng, et al. (2005). "Coupled folding and binding with alpha-helix-forming molecular recognition elements." Biochemistry **44**(37): 12454-70.
Oostenbrink, C., T. A. Soares, et al. (2005). "Validation of the 53A6 GROMOS force field." Eur Biophys J **34**(4): 273-84.
Oostenbrink, C., A. Villa, et al. (2004). "A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6." J Comput Chem **25**(13): 1656-76.
Ozduman, K., G. Wollmann, et al. (2008). "Systemic vesicular stomatitis virus selectively destroys multifocal glioma and metastatic carcinoma in brain." Journal of Neuroscience **28**(8): 1882-1882.
Pace, C. N., B. M. P. Huyghues-Despointes, et al. (2010). "Urea denatured state ensembles contain extensive secondary structure that is increased in hydrophobic proteins." Protein Science **19**(5): 929-943.
Pak, C. C., A. Puri, et al. (1997). "Conformational changes and fusion activity of vesicular stomatitis virus glycoprotein: [125I]iodonaphthyl azide photolabeling studies in biological membranes." Biochemistry **36**(29): 8890-6.
Pal, R., Y. Barenholz, et al. (1987). "Vesicular stomatitis virus membrane proteins and their interactions with lipid bilayers." Biochim Biophys Acta **906**(2): 175-93.
Palin, A., A. Chattopadhyay, et al. (2007). "An optimized vaccine vector based on recombinant vesicular stomatitis virus gives high-level, long-term protection against Yersinia pestis challenge." Vaccine **25**(4): 741-750.
Palma, P. N., L. Krippahl, et al. (2000). "BiGGER: a new (soft) docking algorithm for predicting protein interactions." Proteins **39**(4): 372-84.
Pappu, R. V. and G. D. Rose (2002). "A simple model for polyproline II structure in unfolded states of alanine-based peptides." Protein Sci **11**(10): 2437-55.
Parrinello, M. and A. Rahman (1981). "Polymorphic transitions in single crystals: A new molecular dynamics method." J. Appl. Phys. **52**: 7182.
Pasdeloup, D., N. Poisson, et al. (2005). "Nucleocytoplasmic shuttling of the rabies virus P protein requires a nuclear localization signal and a CRM1-dependent nuclear export signal." Virology **334**(2): 284-93.
Pasdeloup, D., N. Poisson, et al. (2005). "Nucleocytoplasmic shuttling of the rabies virus P protein requires a nuclear localization signal and a CRM1-dependent nuclear export signal." Virology **334**(2): 284-293.
Pasteur, L. (1886). Résultats de l'application de la méthode pour prévenir la rage après morsure, Gauthier-Villars.
Pastoret, P. P. and B. Brochier (2009). "The development and use of a vaccinia-rabies recombinant oral vaccine for the control of wildlife rabies; a link between Jenner and Pasteur." Epidemiology and infection **116**(03): 235-240.
Pattnaik, A. K., L. Hwang, et al. (1997). "Phosphorylation within the amino-terminal acidic domain I of the phosphoprotein of vesicular stomatitis virus is required for transcription but not for replication." J Virol **71**(11): 8167-75.
Pattnaik, A. K. and G. W. Wertz (1990). "Replication and amplification of defective interfering particle RNAs of vesicular stomatitis virus in cells expressing viral proteins from vectors containing cloned cDNAs." J Virol **64**(6): 2948-57.
Patton, J. T., N. L. Davis, et al. (1984). "N protein alone satisfies the requirement for protein synthesis during RNA replication of vesicular stomatitis virus." J Virol **49**(2): 303-9.
Paul, P. R., D. Chattopadhyay, et al. (1988). "The functional domains of the phosphoprotein (NS) of vesicular stomatitis virus (Indiana serotype)." Virology **166**(2): 350-7.
Pedretti, A., L. Villa, et al. (2004). "VEGA--an open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming." J Comput Aided Mol Des **18**(3): 167-73.
Peeters, B. P. H., Y. K. Gruijthuijsen, et al. (2000). "Genome replication of Newcastle disease virus: involvement of the rule-of-six." Archives of virology **145**(9): 1829-1845.
Pei, J. and N. V. Grishin (2001). "AL2CO: calculation of positional conservation in a protein sequence alignment." Bioinformatics **17**(8): 700-12.
Pelet, T., C. Delenda, et al. (1996). "Partial characterization of a Sendai virus replication promoter and the rule of six." Virology **224**(2): 405-414.

Peluso, R. W. (1988). "Kinetic, quantitative, and functional analysis of multiple forms of the vesicular stomatitis virus nucleocapsid protein in infected cells." J Virol **62**(8): 2799-807.
Peluso, R. W. and S. A. Moyer (1983). "Initiation and replication of vesicular stomatitis virus genome RNA in a cell-free system." Proc Natl Acad Sci U S A **80**(11): 3198-202.
Peluso, R. W. and S. A. Moyer (1988). "Viral proteins required for the in vitro replication of vesicular stomatitis virus defective interfering particle genome RNA." Virology **162**(2): 369-76.
Perlman, S. M. and A. S. Huang (1973). "RNA synthesis of vesicular stomatitis virus V. Interactions between transcription and replication." Journal of Virology **12**(6): 1395-1395.
Permyakov, S. E., I. S. Millett, et al. (2003). "Natively unfolded C-terminal domain of caldesmon remains substantially unstructured after the effective binding to calmodulin." Proteins **53**(4): 855-62.
Perrault, J., G. M. Clinton, et al. (1983). "RNP template of vesicular stomatitis virus regulates transcription and replication functions." Cell **35**(1): 175-185.
Petersen, J. M., L. S. Her, et al. (2000). "The matrix protein of vesicular stomatitis virus inhibits nucleocytoplasmic transport when it is in the nucleus and associated with nuclear pore complexes." Mol Cell Biol **20**(22): 8590-601.
Petoukhov, M. V. and D. I. Svergun (2005). "Global rigid body modeling of macromolecular complexes against small-angle scattering data." Biophys J **89**(2): 1237-50.
Phillips, R. J., A. C. R. Samson, et al. (1998). "Nucleotide sequence of the 5 -terminus of Newcastle disease virus and assembly of the complete genomic sequence: agreement with the "rule of six"." Archives of virology **143**(10): 1993-2002.
Pike, L. J. (2009). "The challenge of lipid rafts." The Journal of Lipid Research **50**(Supplement): S323-S323.
Poch, O., B. M. Blumberg, et al. (1990). "Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains." J Gen Virol **71 (Pt 5)**: 1153-62.
Poch, O., I. Sauvaget, et al. (1989). "Identification of four conserved motifs among the RNA-dependent polymerase encoding elements." Embo J **8**(12): 3867-74.
Poisson, N., E. Real, et al. (2001). "Molecular basis for the interaction between rabies virus phosphoprotein P and the dynein light chain LC8: dissociation of dynein-binding properties and transcriptional functionality of P." J Gen Virol **82**(Pt 11): 2691-6.
Potterton, E., P. Briggs, et al. (2003). "A graphical user interface to the CCP4 program suite." Acta Crystallogr D Biol Crystallogr **59**(Pt 7): 1131-7.
Powell, H. R. (1999). "The Rossmann Fourier autoindexing algorithm in MOSFLM." Acta Crystallogr D Biol Crystallogr **55**(Pt 10): 1690-5.
Pringle, C. R. (1997). "The order Mononegavirales--current status." Arch Virol **142**(11): 2321-6.
Ptashne, M. and A. Gann (1997). "Transcriptional activation by recruitment." Nature **386**(6625): 569-77.
Putnam, C. D., M. Hammel, et al. (2007). "X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution." Q Rev Biophys **40**(3): 191-285.
Qanungo, K. R., D. Shaji, et al. (2004). "Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA." Proceedings of the National Academy of Sciences **101**(16): 5952-5952.
Qanungo, K. R., D. Shaji, et al. (2004). "Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA." Proc Natl Acad Sci U S A **101**(16): 5952-7.
Qu, Y., C. L. Bolen, et al. (1998). "Osmolyte-driven contraction of a random coil protein." Proc Natl Acad Sci U S A **95**(16): 9268-73.
Quiroz, E., N. Moreno, et al. (1988). "A human case of encephalitis associated with vesicular stomatitis virus (Indiana serotype) infection." The American journal of tropical medicine and hygiene **39**(3): 312-312.
Raha, T., D. Chattopadhyay, et al. (1999). "A phosphorylation-induced major structural change in the N-terminal domain of the P protein of Chandipura virus." Biochemistry **38**(7): 2110-6.
Raha, T., E. Samal, et al. (2000). "N-terminal region of P protein of Chandipura virus is responsible for phosphorylation-mediated homodimerization." Protein Eng **13**(6): 437-44.
Rahmeh, A. A., J. Li, et al. (2009). "Ribose 2'-O Methylation of the Vesicular Stomatitis Virus mRNA Cap Precedes and Facilitates Subsequent Guanine-N-7 Methylation by the Large Polymerase Protein." Journal of Virology **83**(21): 11043-11043.
Ramsburg, E., N. F. Rose, et al. (2004). "Highly effective control of an AIDS virus challenge in macaques by using vesicular stomatitis virus and modified vaccinia virus Ankara vaccine vectors in a single-boost protocol." Journal of Virology **78**(8): 3930-3930.
Rao, B. L., A. Basu, et al. (2004). "A large outbreak of acute encephalitis with high fatality rate in children in Andhra Pradesh, India, in 2003, associated with Chandipura virus." The Lancet **364**(9437): 869-874.

Raoult, D., S. Audic, et al. (2004). "The 1.2-megabase genome sequence of Mimivirus." Science **306**(5700): 1344-1344.
Raux, H., A. Flamand, et al. (2000). "Interaction of the rabies virus P protein with the LC8 dynein light chain." J Virol **74**(21): 10212-6.
Raymond, D. D., M. E. Piper, et al. (2010). "Structure of the Rift Valley fever virus nucleocapsid protein reveals another architecture for RNA encapsidation." Proc Natl Acad Sci U S A **107**(26): 11769-74.
Reagan, K. J. and W. H. Wunner (1985). "Rabies virus interaction with various cell lines is independent of the acetylcholine receptor." Arch Virol **84**(3-4): 277-82.
Renesto, P., C. Abergel, et al. (2006). "Mimivirus giant particles incorporate a large fraction of anonymous and unique gene products." Journal of Virology **80**(23): 11678-11678.
Ribeiro, E. A., Jr., A. Favier, et al. (2008). "Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein." J Mol Biol **382**(2): 525-38.
Ribeiro, E. A., Jr., A. Favier, et al. (2008). "Solution Structure of the C-Terminal Nucleoprotein-RNA Binding Domain of the Vesicular Stomatitis Virus Phosphoprotein." J Mol Biol **382**: 525-538.
Ribeiro, E. A., C. Leyrat, et al. (2009). "Binding of rabies virus polymerase cofactor to recombinant circular nucleoprotein-RNA complexes." J Mol Biol **394**(3): 558-75.
Ribeiro, E. D., Jr., C. Leyrat, et al. (2009). "Binding of Rabies Virus Polymerase Cofactor to Recombinant Circular Nucleoprotein-RNA Complexes." J Mol Biol **394**: 558-575.
Ribeiro Ede, A., Jr., C. Leyrat, et al. (2009). "Binding of rabies virus polymerase cofactor to recombinant circular nucleoprotein-RNA complexes." J Mol Biol **394**(3): 558-75.
Rieder, M. and K. K. Conzelmann (2009). "Rhabdovirus evasion of the interferon system." J Interferon Cytokine Res **29**(9): 499-509.
Roche, S., A. A. V. Albertini, et al. (2008). "Structures of vesicular stomatitis virus glycoprotein: membrane fusion revisited." Cellular and molecular life sciences **65**(11): 1716-1728.
Roche, S., S. Bressanelli, et al. (2006). "Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G." Science **313**(5784): 187-91.
Roche, S. and Y. Gaudin (2002). "Characterization of the equilibrium between the native and fusion-inactive conformation of rabies virus glycoprotein indicates that the fusion complex is made of several trimers." Virology **297**(1): 128-35.
Roche, S. and Y. Gaudin (2004). "Evidence that rabies virus forms different kinds of fusion machines with different pH thresholds for fusion." Journal of Virology **78**(16): 8746-8746.
Roche, S., F. A. Rey, et al. (2007). "Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G." Science **315**(5813): 843-8.
Rodriguez, L. L. (2002). "Emergence and re-emergence of vesicular stomatitis in the United States." Virus Res **85**(2): 211-9.
Roediger, E. K., K. Kugathasan, et al. (2008). "Heterologous Boosting of Recombinant Adenoviral Prime Immunization With a Novel Vesicular Stomatitis Virus–vectored Tuberculosis Vaccine." Molecular Therapy **16**(6): 1161-1169.
Roine, R. O., M. Hillrom, et al. (1988). "Fatal encephalitis caused by a bat-borne rabies-related virus: clinical findings." Brain **111**(6): 1505-1505.
Rose, J. K., G. A. Adams, et al. (1984). "The presence of cysteine in the cytoplasmic domain of the vesicular stomatitis virus glycoprotein is required for palmitate addition." Proc Natl Acad Sci U S A **81**(7): 2050-4.
Rose, J. K. and L. Iverson (1979). "Nucleotide sequences from the 3'-ends of vesicular stomatitis virus mRNA's as determined from cloned DNA." J Virol **32**(2): 404-11.
Rose, N. F., J. Publicover, et al. (2008). "Hybrid alphavirus-rhabdovirus propagating replicon particles are versatile and potent vaccine vectors." Proc Natl Acad Sci U S A **105**(15): 5839-43.
Roy, A., A. Kucukural, et al. (2010). "I-TASSER: a unified platform for automated protein structure and function prediction." Nat Protoc **5**(4): 725-38.
Rudolph, M. G., I. Kraus, et al. (2003). "Crystal structure of the borna disease virus nucleoprotein." Structure **11**(10): 1219-26.
Rühle, V. (2008). "Pressure coupling/barostats."
Ruigrok, R. W. H. and T. Crépin (2010). "Nucleoproteins of Negative Strand RNA Viruses; RNA Binding, Oligomerisation and Binding to Polymerase Co-Factor." Viruses **2**(1): 27-32.
Ruigrok, T. J. H., A. Pijpers, et al. (2008). "Multiple cerebellar zones are involved in the control of individual muscles: a retrograde transneuronal tracing study with rabies virus in the rat." European Journal of Neuroscience **28**(1): 181-200.
Rupprecht, C. E., J. S. Smith, et al. (1995). "The ascension of wildlife rabies: a cause for public health concern or intervention?" Emerging infectious diseases **1**(4): 107-107.

Rupprecht, C. E., R. Willoughby, et al. (2006). "Current and future trends in the prevention, treatment and control of rabies." Expert Rev Anti Infect Ther **4**(6): 1021-38.
Sabeta, C. T., W. Markotter, et al. (2007). "Mokola virus in domestic mammals, South Africa."
Sadqi, M., L. J. Lapidus, et al. (2003). "How fast is protein hydrophobic collapse?" Proc Natl Acad Sci U S A **100**(21): 12117-22.
Samal, S. K. and P. L. Collins (1996). "RNA replication by a respiratory syncytial virus RNA analog does not obey the rule of six and retains a nonviral trinucleotide extension at the leader end." J Virol **70**(8): 5075-82.
Sanchez-Puig, N., D. B. Veprintsev, et al. (2005). "Human full-length Securin is a natively unfolded protein." Protein Sci **14**(6): 1410-8.
Schanda, P., H. Van Melckebeke, et al. (2006). "Speeding up three-dimensional protein NMR experiments to a few minutes." J Am Chem Soc **128**(28): 9042-3.
Schlegel, R., T. S. Tralka, et al. (1983). "Inhibition of VSV binding and infectivity by phosphatidylserine: is phosphatidylserine a VSV-binding site?" Cell **32**(2): 639-46.
Schmidt, M. F. and M. J. Schlesinger (1979). "Fatty acid binding to vesicular stomatitis virus glycoprotein: a new type of post-translational modification of the viral glycoprotein." Cell **17**(4): 813-9.
Schnell, M. J., L. Buonocore, et al. (1998). "Requirement for a non-specific glycoprotein cytoplasmic domain sequence to drive efficient budding of vesicular stomatitis virus." EMBO J **17**(5): 1289-96.
Schnell, M. J., L. Buonocore, et al. (1996). "Foreign glycoproteins expressed from recombinant vesicular stomatitis viruses are incorporated efficiently into virus particles." Proceedings of the National Academy of Sciences **93**(21): 11359-11359.
Schnell, M. J., L. Buonocore, et al. (1996). "The minimal conserved transcription stop-start signal promotes stable expression of a foreign gene in vesicular stomatitis virus." J Virol **70**(4): 2318-23.
Schnell, M. J., J. P. McGettigan, et al. (2010). "The cell biology of rabies virus: using stealth to reach the brain." Nat Rev Microbiol **8**(1): 51-61.
Schoehn, G., F. Iseni, et al. (2001). "Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site." J Virol **75**(1): 490-8.
Schoehn, G., F. Iseni, et al. (2001). "Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site." Journal of Virology **75**(1): 490-490.
Schoehn, G., M. Mavrakis, et al. (2004). "The 12 A structure of trypsin-treated measles virus N-RNA." J Mol Biol **339**(2): 301-12.
Schubert, M., G. G. Harmison, et al. (1984). "Primary structure of the vesicular stomatitis virus polymerase (L) gene: evidence for a high frequency of mutations." J Virol **51**(2): 505-14.
Schwartz, J. A., L. Buonocore, et al. (2007). "Vesicular stomatitis virus vectors expressing avian influenza H5 HA induce cross-neutralizing antibodies and long-term protection." Virology **366**(1): 166-173.
Schwemmle, M., B. De, et al. (1997). "Borna disease virus P-protein is phosphorylated by protein kinase Cepsilon and casein kinase II." J Biol Chem **272**(35): 21818-23.
Schymkowitz, J., J. Borg, et al. (2005). "The FoldX web server: an online force field." Nucleic Acids Res **33**(Web Server issue): W382-8.
Seganti, L., F. Superti, et al. (1986). "Study of receptors for vesicular stomatitis virus in vertebrate and invertebrate cells." Microbiologica **9**(3): 259-67.
Sellers, R. F. and A. R. Maarouf (1990). "Trajectory analysis of winds and vesicular stomatitis in North America, 1982-5." Epidemiol Infect **104**(2): 313-28.
Semenyuk, A. V. and D. Svergun (1991). "GNOM - a program package for small-angle scattering data processing." J. Appl. Crystallog. **24**: 537-540.
Sha, B. and M. Luo (1997). "Structure of a bifunctional membrane-RNA binding protein, influenza virus matrix protein M1." Nat Struct Biol **4**(3): 239-44.
Shatkin, A. J. and J. D. Sipe (1968). "RNA polymerase activity in purified reoviruses." Proc Natl Acad Sci U S A **61**(4): 1462-9.
Shayam, C., A. K. Duggal, et al. (2006). "Post-exposure prophylaxis for rabies." J Indian Acad Clin Med **7**(1): 39-46.
Shen, Y. and A. Bax (2007). "Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology." J Biomol NMR **38**(4): 289-302.
Shi, W., Q. Tang, et al. (2009). "Antitumor and antimetastatic activities of vesicular stomatitis virus matrix protein in a murine model of breast cancer." Journal of molecular medicine **87**(5): 493-506.
Shimizu, K., N. Ito, et al. (2006). "Sensitivity of rabies virus to type I interferon is determined by the phosphoprotein gene." Microbiol Immunol **50**(12): 975-8.
Shoemaker, B. A., J. J. Portman, et al. (2000). "Speeding molecular recognition by using the folding funnel: the fly-casting mechanism." Proc Natl Acad Sci U S A **97**(16): 8868-73.

Simpson, R. W. and R. E. Hauser (1966). "Structural components of vesicular stomatitis virus." Virology **29**(4): 654-667.
Sleat, D. E. and A. K. Banerjee (1993). "Transcriptional activity and mutational analysis of recombinant vesicular stomatitis virus RNA polymerase." J Virol **67**(3): 1334-9.
Sokol, F., H. D. Schlumberger, et al. (1969). "Biochemical and biophysical studies on the nucleocapsid and on the RNA of rabies virus." Virology **38**(4): 651-665.
Soria, M., S. P. Little, et al. (1974). "Characterization of vesicular stomatitis virus nucleocapsids. I. Complementary 40 S RNA molecules in nucleocapsids." Virology **61**(1): 270-80.
Spadafora, D., D. M. Canter, et al. (1996). "Constitutive phosphorylation of the vesicular stomatitis virus P protein modulates polymerase complex formation but is not essential for transcription or replication." J Virol **70**(7): 4538-48.
Spera, S. and A. Bax (1991). "Empirical Correlation between Protein Backbone Conformation and Cα and Cβ ^{13}C Nuclear Magnetic Resonance Chemical Shifts." J Am Chem Soc **113**: 5490-5492.
Spiropoulou, C. F. and S. T. Nichol (1993). "A small highly basic protein is encoded in overlapping frame within the P gene of vesicular stomatitis virus." Journal of Virology **67**(6): 3103-3103.
Spolar, R. S. and M. T. Record, Jr. (1994). "Coupling of local folding to site-specific binding of proteins to DNA." Science **263**(5148): 777-84.
Srinivasan, A., E. C. Burton, et al. (2005). "Transmission of rabies virus from an organ donor to four transplant recipients." The New England journal of medicine **352**(11): 1103-1103.
Steele, J. H. and G. M. Baer (1975). The natural history of rabies, Academic Press.
Stephens, E. B. and R. W. Compans (1988). "Assembly of animal viruses at cellular membranes." Annu Rev Microbiol **42**: 489-516.
Stillman, E. A. and M. Whitt (1999). "Transcript initiation and 5'-end modifications are separable events during vesicular stomatitis virus transcription." J. Virol. **73**: 7199-7209.
Stillman, E. A. and M. A. Whitt (1998). "The length and sequence composition of vesicular stomatitis virus intergenic regions affect mRNA levels and the site of transcript initiation." J Virol **72**(7): 5565-72.
Stillman, E. A. and M. A. Whitt (1999). "Transcript initiation and 5'-end modifications are separable events during vesicular stomatitis virus transcription." Journal of Virology **73**(9): 7199-7199.
Streicker, D. G., A. S. Turmelle, et al. (2010). "Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats." Science **329**(5992): 676-9.
Stricker, R., G. Mottet, et al. (1994). "The Sendai virus matrix protein appears to be recruited in the cytoplasm by the viral nucleocapsid to function in viral assembly and budding." J Gen Virol **75 (Pt 5)**: 1031-42.
Strous, G. J., R. Willemsen, et al. (1983). "Vesicular stomatitis virus glycoprotein, albumin, and transferrin are transported to the cell surface via the same Golgi vesicles." The Journal of Cell Biology **97**(6): 1815-1815.
Sugase, K., H. J. Dyson, et al. (2007). "Mechanism of coupled folding and binding of an intrinsically disordered protein." Nature **447**: 920-921.
Summers, W. C. and R. B. Siegel (1970). "Transcription of late phage RNA by T7 RNA polymerase."
Sun, X., S. L. Roth, et al. (2010). "Internalization and fusion mechanism of vesicular stomatitis virus and related rhabdoviruses." Future Virology **5**(1): 85-96.
Sun, X., V. K. Yau, et al. (2005). "Role of clathrin-mediated endocytosis during vesicular stomatitis virus entry into host cells." Virology **338**(1): 53-60.
Superti, F., C. Girmenta, et al. (1986). "Role of sialic acid in cell receptors for vesicular stomatitis virus." Acta Virol **30**(1): 10-8.
Superti, F., B. Hauttecoeur, et al. (1986). "Involvement of gangliosides in rabies virus infection." J Gen Virol **67 (Pt 1)**: 47-56.
Superti, F., L. Seganti, et al. (1984). "Role of phospholipids in rhabdovirus attachment to CER cells. Brief report." Arch Virol **81**(3-4): 321-8.
Svergun, D., C. Barberato, et al. (1995). "CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates." J. Appl. Cryst. **28**: 768-773.
Svergun, D. I. (1999). "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing." Biophys J **76**(6): 2879-86.
Svergun, D. I., M. V. Petoukhov, et al. (2001). "Determination of domain structure of proteins from X-ray solution scattering." Biophys J **80**(6): 2946-53.
Swinteck, B. D. and D. S. Lyles (2008). "Plasma membrane microdomains containing vesicular stomatitis virus M protein are separate from microdomains containing G protein and nucleocapsids." J Virol **82**(11): 5536-47.
Takacs, A. M. and A. K. Banerjee (1995). "Efficient interaction of the vesicular stomatitis virus P protein with the L protein or the N protein in cells expressing the recombinant proteins." Virology **208**(2): 821-6.

Takacs, A. M., S. Barik, et al. (1992). "Phosphorylation of specific serine residues within the acidic domain of the phosphoprotein of vesicular stomatitis virus regulates transcription in vitro." J Virol **66**(10): 5842-8.
Takacs, A. M., T. Das, et al. (1993). "Mapping of interacting domains between the nucleocapsid protein and the phosphoprotein of vesicular stomatitis virus by using a two-hybrid system." Proc Natl Acad Sci U S A **90**(21): 10375-9.
Takacs, A. M., K. G. Perrine, et al. (1991). "Alteration of specific amino acid residues in the acidic domain I of VSV phosphoprotein (P) converts a GAL4-P(I) hybrid into a transcriptional activator." New Biol **3**(6): 581-91.
Tan, G. S., M. A. Preuss, et al. (2007). "The dynein light chain 8 binding motif of rabies virus phosphoprotein promotes efficient viral transcription." Proc Natl Acad Sci U S A **104**(17): 7229-34.
Tanford, C. (1961). Physical chemistry of macromolecules. New York.
Tanford, C. (1968). "Protein denaturation." Adv Protein Chem **23**: 121-282.
Tang, Y., O. Rampin, et al. (1999). "Spinal and brain circuits to motoneurons of the bulbospongiosus muscle: retrograde transneuronal tracing with rabies virus." The Journal of Comparative Neurology **414**(2): 167-192.
Tarbouriech, N., J. Curran, et al. (2000). "On the domain structure and the polymerization state of the sendai virus P protein." Virology **266**(1): 99-109.
Tarbouriech, N., J. Curran, et al. (2000). "Tetrameric coiled coil domain of Sendai virus phosphoprotein." Nat Struct Biol **7**(9): 777-81.
Tawar, R. G., S. Duquerroy, et al. (2009). "Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus." Science **326**(5957): 1279-83.
Tesh, R., S. Saidi, et al. (1977). "Isfahan virus, a new vesiculovirus infecting humans, gerbils, and sandflies in Iran." Am J Trop Med Hyg **26**(2): 299-306.
Tesh, R., S. Saidi, et al. (1977). "Studies on the epidemiology of sandfly fever in Iran. I. Virus isolates obtained from Phlebotomus." Am J Trop Med Hyg **26**(2): 282-7.
Tesh, R. B., J. Boshell, et al. (1987). "Natural infection of humans, animals, and phlebotomine sand flies with the Alagoas serotype of vesicular stomatitis virus in Colombia." The American journal of tropical medicine and hygiene **36**(3): 653-653.
Tesh, R. B., B. N. Chaniotis, et al. (1972). "Vesicular stomatitis virus (Indiana serotype): transovarial transmission by phlebotomine sandflies." Science **175**(29): 1477-9.
Testa, D. and A. K. Banerjee (1977). "Two methyltransferase activities in the purified virions of vesicular stomatitis virus." Journal of Virology **24**(3): 786-786.
Théodoridès, J. (1986). Histoire de la rage. Paris, Masson.
Thomas, D., W. W. Newcomb, et al. (1985). "Mass and molecular composition of vesicular stomatitis virus: a scanning transmission electron microscopy analysis." J Virol **54**(2): 598-607.
Thornton, G. B., B. P. De, et al. (1984). "Interaction of L and NS proteins of vesicular stomatitis virus with its template ribonucleoprotein during RNA synthesis in vitro." J Gen Virol **65 (Pt 3)**: 663-8.
Thoulouze, M. I., E. Bouguyon, et al. (2004). "Essential role of the NV protein of Novirhabdovirus for pathogenicity in rainbow trout." J Virol **78**(8): 4098-107.
Thoulouze, M. I., M. Lafage, et al. (1998). "The neural cell adhesion molecule is a receptor for rabies virus." J Virol **72**(9): 7181-90.
Timmins, J., R. W. Ruigrok, et al. (2004). "Structural studies on the Ebola virus matrix protein VP40 indicate that matrix proteins of enveloped RNA viruses are analogues but not homologues." FEMS Microbiol Lett **233**(2): 179-86.
Tina, K. G., R. Bhadra, et al. (2007). "PIC: Protein Interactions Calculator." Nucleic Acids Res **35**(Web Server issue): W473-6.
Tirawatnpong, S., T. Hemachudha, et al. (1989). "Regional distribution of rabies viral antigen in central nervous system of human encephalitic and paralytic rabies." Journal of the neurological sciences **92**(1): 91-99.
Tokuriki, N., C. J. Oldfield, et al. (2009). "Do viral proteins possess unique biophysical features?" Trends Biochem Sci **34**(2): 53-9.
Tompa, P. (2002). "Intrinsically unstructured proteins." Trends Biochem Sci **27**(10): 527-33.
Tompa, P. and M. Fuxreiter (2008). "Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions." Trends Biochem Sci **33**(1): 2-8.
Tompa, P., C. Szasz, et al. (2005). "Structural disorder throws new light on moonlighting." Trends Biochem Sci **30**(9): 484-9.
Tordo, N. and A. Kouknetzoff (1993). "The rabies virus genome: an overview." Onderstepoort J Vet Res **60**(4): 263-9.
Tordo, N., O. Poch, et al. (1986). "Primary structure of leader RNA and nucleoprotein genes of the rabies genome: segmented homology with VSV." Nucleic Acids Res **14**(6): 2671-83.

Tordo, N., O. Poch, et al. (1986). "Walking along the rabies genome: is the large G-L intergenic region a remnant gene?" Proc Natl Acad Sci U S A **83**(11): 3914-8.
Tordo, N., O. Poch, et al. (1988). "Completion of the rabies virus genome sequence determination: highly conserved domains among the L (polymerase) proteins of unsegmented negative-strand RNA viruses." Virology **165**(2): 565-76.
Toriumi, H., Y. Eriguchi, et al. (2004). "Further studies on the hyperphosphorylated form (p40) of the rabies virus nominal phosphoprotein (P)." Microbiol Immunol **48**(11): 865-74.
Toriumi, H., Y. Honda, et al. (2002). "Structural relationship between nucleocapsid-binding activity of the rabies virus phosphoprotein (P) and exposure of epitope 402-13 located at the C terminus." J Gen Virol **83**(Pt 12): 3035-43.
Toriumi, H. and A. Kawai (2004). "Association of rabies virus nominal phosphoprotein (P) with viral nucleocapsid (NC) is enhanced by phosphorylation of the viral nucleoprotein (N)." Microbiol Immunol **48**(5): 399-409.
Torreira, E., G. Schoehn, et al. (2007). "Three-dimensional model for the isolated recombinant influenza virus polymerase heterotrimer." Nucleic Acids Res **35**(11): 3774-83.
Tosatto, S. C., E. Bindewald, et al. (2002). "A divide and conquer approach to fast loop modeling." Protein Eng **15**(4): 279-86.
Tsai, C. J., S. Kumar, et al. (1999). "Folding funnels, binding funnels, and protein function." Protein Sci **8**(6): 1181-90.
Tuffereau, C., J. Benejean, et al. (1998). "Low-affinity nerve-growth factor receptor (P75NTR) can serve as a receptor for rabies virus." Embo J **17**(24): 7250-9.
Tuffereau, C., E. Desmezieres, et al. (2001). "Interaction of lyssaviruses with the low-affinity nerve-growth factor receptor p75NTR." J Gen Virol **82**(Pt 12): 2861-7.
Turjanski, A. G., J. S. Gutkind, et al. (2008). "Binding-induced folding of a natively unstructured transcription factor." PLoS Comput Biol **4**(4): e1000060.
Ugolini, G. (1995). "Specificity of rabies virus as a transneuronal tracer of motor networks: transfer from hypoglossal motoneurons to connected second-order and higher order central nervous system cell groups." The Journal of Comparative Neurology **356**(3): 457-480.
Uversky, V. N. (1993). "Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule." Biochemistry **32**(48): 13288-98.
Uversky, V. N. (2002). "Natively unfolded proteins: a point where biology waits for physics." Protein Sci **11**(4): 739-56.
Uversky, V. N. (2009). "Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding." The protein journal **28**(7): 305-325.
Uversky, V. N., J. R. Gillespie, et al. (2000). "Why are "natively unfolded" proteins unstructured under physiologic conditions?" Proteins **41**(3): 415-27.
Uversky, V. N., J. R. Gillespie, et al. (1999). "Natively unfolded human prothymosin alpha adopts partially folded collapsed conformation at acidic pH." Biochemistry **38**(45): 15009-16.
Uversky, V. N., J. Li, et al. (2001). "Evidence for a partially folded intermediate in alpha-synuclein fibril formation." J Biol Chem **276**(14): 10737-44.
Uversky, V. N., J. Li, et al. (2001). "Trimethylamine-N-oxide-induced folding of alpha-synuclein." FEBS Lett **509**(1): 31-5.
Uversky, V. N., J. Li, et al. (2002). "Biophysical properties of the synucleins and their propensities to fibrillate: inhibition of alpha-synuclein assembly by beta- and gamma-synucleins." J Biol Chem **277**(14): 11970-8.
Uversky, V. N., C. J. Oldfield, et al. (2005). "Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling." J Mol Recognit **18**(5): 343-84.
Vacic, V., C. J. Oldfield, et al. (2007). "Characterization of molecular recognition features, MoRFs, and their binding partners." J Proteome Res **6**(6): 2351-66.
Vajda, S. and D. Kozakov (2009). "Convergence and combination of methods in protein-protein docking." Current opinion in structural biology **19**(2): 164-170.
van der Spoel, D., E. Lindahl, et al. (2005). "GROMACS: Fast, Flexible and Free." J. Comp. Chem. **26**: 1701-1718.
van Thiel, P., R. M. A. de Bie, et al. (2009). "Fatal Human Rabies due to Duvenhage Virus from a Bat in Kenya: Failure of Treatment with Coma-Induction, Ketamine, and Antiviral Drugs."
Vidal, S. and D. Kolakofsky (1989). "Modified model for the switch from Sendai virus transcription to replication." J Virol **63**(5): 1951-8.
Vidy, A., J. Bougrini, et al. (2007). "The nulceocytoplasmic rabies P protein counteracts interferons signaling by inhibiting both nulear accumulation and DNA binding of STAT1." J Virol. **81**: 4255-4263.

Vidy, A., M. Chelbi-Alix, et al. (2005). "Rabies virus P protein interacts with STAT1 and inhibits interferon signal transduction pathways." J Virol **79**(22): 14411-20.
Vidy, A., J. El Bougrini, et al. (2007). "The nucleocytoplasmic rabies virus P protein counteracts interferon signaling by inhibiting both nuclear accumulation and DNA binding of STAT1." Journal of Virology **81**(8): 4255-4255.
Vidy, A., J. El Bougrini, et al. (2007). "The nucleocytoplasmic rabies virus P protein counteracts interferon signaling by inhibiting both nuclear accumulation and DNA binding of STAT1." J Virol **81**(8): 4255-63.
Volkov, V. V. and D. I. Svergun (2003). "Uniqueness of ab initio shape determination in small-angle scattering." J. Appl. Cryst. **36**: 860-864.
von Kobbe, C., J. van Deursen, et al. (2000). "Vesicular stomatitis virus matrix protein inhibits host cell gene expression by targeting the nucleoporin Nup98." Molecular cell **6**(5): 1243-1252.
Vulliemoz, D., S. Cordey, et al. (2005). "Nature of a paramyxovirus replication promoter influences a nearby transcription signal." Journal of General Virology **86**(1): 171-171.
Vulliemoz, D. and L. Roux (2001). "" Rule of Six": How Does the Sendai Virus RNA Polymerase Keep Count?" Journal of Virology **75**(10): 4506-4506.
Wagner, R. R. and J. K. Rose (1996). Rhabdoviridae: The viruses and their replication. Fields Virology, 3rd edn. B. N. Fields, D. M. Knipe, P. M. Howleyet al. New York, Lippincott-Raven Publishers**:** 1121-1135.
Wagner, R. R., T. C. Schnaitman, et al. (1969). "Protein composition of the structural components of vesicular stomatitis virus." Journal of Virology **3**(6): 611-611.
Wang, C., P. Bradley, et al. (2007). "Protein-protein docking with backbone flexibility." J Mol Biol **373**(2): 503-19.
Wang, C., O. Schueler-Furman, et al. (2007). "RosettaDock in CAPRI rounds 6-12." Proteins **69**(4): 758-63.
Wang, Z. W., L. Sarmento, et al. (2005). "Attenuated rabies virus activates, while pathogenic rabies virus evades, the host innate immune responses in the central nervous system." J Virol **79**(19): 12554-65.
Ward, J. J., J. S. Sodhi, et al. (2004). "Prediction and functional analysis of native disorder in proteins from the three kingdoms of life." J Mol Biol **337**(3): 635-45.
Warrell, M. J. and D. A. Warrell (2004). "Rabies and other lyssavirus diseases." Lancet **363**(9413): 959-69.
Webster, L. T. and A. D. Clow (1936). "Propagation of rabies virus in tissue culture and the successful use of culture virus as an antirabic vaccine." Science **84**: 487-488.
Weissenhorn, W., A. Dessen, et al. (1999). "Structural basis for membrane fusion by enveloped viruses." Molecular membrane biology **16**(1): 3-9.
Wertz, G. W., V. P. Perepelitsa, et al. (1998). "Gene rearrangement attenuates expression and lethality of a nonsegmented negative strand RNA virus." Proc Natl Acad Sci U S A **95**(7): 3501-6.
Wertz, G. W., S. Whelan, et al. (1994). "Extent of terminal complementarity modulates the balance between transcription and replication of vesicular stomatitis virus RNA." Proceedings of the National Academy of Sciences of the United States of America **91**(18): 8587-8587.
Weyer, J., I. V. Kuzmin, et al. (2008). "Cross-protective and cross-reactive immune responses to recombinant vaccinia viruses expressing full-length lyssavirus glycoprotein genes." Epidemiol Infect **136**(5): 670-8.
Whelan, S. P., J. N. Barr, et al. (2004). "Transcription and replication of nonsegmented negative-strand RNA viruses." Curr Top Microbiol Immunol **283**: 61-119.
Whelan, S. P. and G. W. Wertz (1999). "The 5' terminal trailer region of vesicular stomatitis virus contains a position-dependent cis-acting signal for assembly of RNA into infectious particles." J Virol **73**(1): 307-15.
Whelan, S. P. and G. W. Wertz (1999). "Regulation of RNA synthesis by the genomic termini of vesicular stomatitis virus: identification of distinct sequences essential for transcription but not replication." J Virol **73**(1): 297-306.
Whelan, S. P. and G. W. Wertz (2002). "Transcription and replication initiate at separate sites on the vesicular stomatitis virus genome." Proc Natl Acad Sci U S A **99**(14): 9178-83.
White, J., M. Kielian, et al. (2009). "Membrane fusion proteins of enveloped animal viruses." Quarterly Reviews of Biophysics **16**(02): 151-195.
White, J., K. Matlin, et al. (1981). "Cell fusion by Semliki Forest, influenza, and vesicular stomatitis viruses." The Journal of cell biology **89**(3): 674-674.
Whitt, M. A., L. Chong, et al. (1989). "Glycoprotein cytoplasmic domain sequences required for rescue of a vesicular stomatitis virus glycoprotein mutant." Journal of Virology **63**(9): 3569-3569.
Whittington, S. J., B. W. Chellgren, et al. (2005). "Urea promotes polyproline II helix formation: implications for protein denatured states." Biochemistry **44**(16): 6269-75.
Wilde, H. (1999). "Failures of Post-Exposure Rabies Prophylaxis." Clin Infect Dis **28**: 143-4.
Wilks, C. R. and J. A. House (2009). "Susceptibility of various animals to the vesiculovirus Piry." Epidemiology and infection **93**(01): 147-155.

Wilks, C. R. and J. A. House (2009). "Susceptibility of various animals to the vesiculoviruses Isfahan and Chandipura." Epidemiology and infection **97**(02): 359-368.
Willoughby Jr, R. E., K. S. Tieves, et al. (2005). "Survival after treatment of rabies with induction of coma." The New England journal of medicine **352**(24): 2508-2508.
Wirblich, C., G. S. Tan, et al. (2008). "PPEY motif within the rabies virus (RV) matrix protein is essential for efficient virion release and RV pathogenicity." J Virol **82**(19): 9730-8.
Wishart, D. S., B. D. Sykes, et al. (1991). "Relationship between nuclear magnetic resonance chemical shift and protein secondary structure." J Mol Biol **222**(2): 311-33.
Wright, P. E. and H. J. Dyson (1999). "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm." J Mol Biol **293**(2): 321-31.
Wright, P. E. and H. J. Dyson (2009). "Linking folding and binding." Curr Opin Struct Biol **19**(1): 31-8.
Wu, S., J. Skolnick, et al. (2007). "Ab initio modeling of small proteins by iterative TASSER simulations." BMC Biol **5**: 17.
Wu, S. and Y. Zhang (2007). "LOMETS: a local meta-threading-server for protein structure prediction." Nucleic Acids Res **35**(10): 3375-82.
Wu, X., X. Lei, et al. (2003). "Rabies virus nucleoprotein is phosphorylated by cellular casein kinase II." Biochem Biophys Res Commun **304**(2): 333-8.
Wu, Y., X. Tian, et al. (2005). "Folding of small helical proteins assisted by small-angle X-ray scattering profiles." Structure **13**(11): 1587-1597.
Wunner, W. H., K. J. Reagan, et al. (1984). "Characterization of saturable binding sites for rabies virus." J Virol **50**(3): 691-7.
Wyatt, P. J. (1998). "Submicrometer Particle Sizing by Multiangle Light Scattering following Fractionation." J Colloid Interface Sci **197**(1): 9-20.
Yang, J., D. C. Hooper, et al. (1998). "The specificity of rabies virus RNA encapsidation by nucleoprotein." Virology **242**(1): 107-17.
Yang, J., H. Koprowski, et al. (1999). "Phosphorylation of rabies virus nucleoprotein regulates viral RNA transcription and replication by modulating leader RNA encapsidation." J Virol **73**(2): 1661-4.
Yap, T. L., T. Xu, et al. (2007). "Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution." J Virol **81**(9): 4753-65.
Ye, Q., R. M. Krug, et al. (2006). "The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA." Nature **444**(7122): 1078-82.
Yilma, T., R. G. Breeze, et al. (1985). "Immune responses of cattle and mice to the G glycoprotein of vesicular stomatitis virus." Advances in experimental medicine and biology (USA).
Young, C. C. and T. F. Sellers (1927). "Laboratory: a new method for staining Negri bodies of rabies." American Journal of Public Health **17**(10): 1080-1080.
Zacharias, M. (2010). "Accounting for conformational changes during protein-protein docking." Current opinion in structural biology **20**(2): 180-186.
Zafrullah, M., M. H. Ozdener, et al. (1997). "The ORF3 protein of hepatitis E virus is a phosphoprotein that associates with the cytoskeleton." Journal of Virology **71**(12): 9045-9045.
Zagrovic, B., J. Lipfert, et al. (2005). "Unusual compactness of a polyproline type II structure." Proc Natl Acad Sci U S A **102**(33): 11698-703.
Zhang, X., T. J. Green, et al. (2008). "Role of intermolecular interactions of vesicular stomatitis virus nucleoprotein in RNA encapsidation." J Virol **82**(2): 674-82.
Zhang, Y. (2008). "I-TASSER server for protein 3D structure prediction." BMC Bioinformatics **9**: 40.
Zhang, Y. (2008). "Progress and challenges in protein structure prediction." Current opinion in structural biology **18**(3): 342-348.
Zheng, W. and S. Doniach (2002). "Protein structure prediction constrained by solution X-ray scattering data and structural homology identification1." Journal of Molecular Biology **316**(1): 173-187.
Zheng, W. and S. Doniach (2005). "Fold recognition aided by constraints from small angle X-ray scattering data." Protein Engineering Design and Selection **18**(5): 209-209.
Zinke, G. G. (1804). Neue Ansichten der Hundswuth; ihrer Ursachen und Folgen, nebst einer sichern Behandlungsart der von tollen Thieren gebissenen Mensch, CE Gabler.

ANNEXES

“Il est pas très sociable ce rat, il a été trouvé dans un caisson de basses…”

(Gabriel Charles Ernest Bouaniche, à propos du rat dans la cuisine)

“Where does the energy come from?”

(Anonymous, NSV Meeting 2010, Brugge)

gcq#21: "What Kind Of Guru are You, Anyway ?"

(F. Zappa)

“The Psycho-Social, Chemical, Biological, And Electro-Magnetic Manipulation Of Human Consciousness”

(Jedi Mind Tricks)

I. REVIEW II: Structural Disorder in Proteins of the Rhabdoviridae Replication Complex

Auteurs: Cédric Leyrat, Francine Gérard, Euripedes de Almeida Ribeiro Jr, Ivan Ivanov, Rob W. H. Ruigrok, and Marc Jamin.

Publié en 2010 dans *Protein & Peptide Letters*, Volume 17, Pages 979-987

Structural Disorder in Proteins of the Rhabdoviridae Replication Complex

Cédric Leyrat, Francine C.A. Gérard, Euripedes de Almeida Ribeiro Jr., Ivan Ivanov, Rob W.H. Ruigrok and Marc Jamin*

Unit of Virus Host Cell Interactions, UMI 3265 UJF-EMBL-CNRS, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

Abstract: *Rhabdoviridae* are single stranded negative sense RNA viruses. The viral RNA condensed by the nucleoprotein (N), the phosphoprotein (P) and the large subunit (L) of the RNA-dependent RNA polymerase are the viral components of the transcription/replication machineries. Both P and N contain intrinsically disordered regions (IDRs) that play different roles in the virus life cycle. Here, we describe the modular organization of P based on recent structural, biophysical and bioinformatics data. We show how flexible loops in N participate in the attachment of P to the N-RNA template by an induced-fit mechanism. Finally, we discuss the roles of IDRs in the mechanism of replication/transcription, and propose a new model for the interaction of the L subunit with its N-RNA template.

Keywords: Rhabdoviridae, replication complex, phosphoprotein, nucleoprotein, intrinsically disordered regions.

THE RHABDOVIRIDAE

Rabies virus (RV) and vesicular stomatitis virus (VSV) are prototypic members of the large family *Rhabdoviridae* that includes numerous pathogens of plants and animals, including man, and are grouped in the *Mononegavirales* (*MNV*) order with the *Filoviridae* (Ebola and Marburg viruses), the *Paramyxoviridae* (measles, mumps, respiratory syncytial viruses) and the *Bornaviridae* (Borna disease virus). *Rhabdoviridae* are enveloped viruses with an overall bullet or bacillus shape, whose genome is made of a linear single-stranded molecule of negative sense RNA. They vary widely in terms of cellular interactions and replicate either in the cytoplasm or in the nucleus, but they share similar genome and structural organizations as well as similar modes of RNA replication and transcription [1, 2]. The viral genome (9 to 18 kb) of the *Rhabdoviridae* comprises up to 10 genes, but both RV and VSV genomes contain only the five genes that are common to all members of the *MNV* (Fig. **1a**). They encode successively from the 3' terminus, the nucleoprotein (N), the phosphoprotein (P), the matrix protein (M), the glycoprotein (G) and the large subunit of the RNA polymerase (L). Only three of these viral proteins, N, P and L, are required to synthesize viral RNA in an efficient and regulated manner [3, 4].

The viral genomic RNA of *Rhabdoviridae* is always encapsidated by N (Fig. **1b**) forming a helical ribonucleocapsid (NC) in which each N protomer binds to nine nucleotides. This N-RNA complex serves as the active template for both transcription and replication by the two-subunit P-L RNA-dependent RNA polymerase [5]. When expressed in bacteria or insect cells, recombinant N protein binds to cellular RNAs and forms circular N_m-RNA complexes in addition to long NCs [6-8] (the subscript m denotes the number of N protomers *per* ring). Circular N_m-RNA complexes containing different numbers of N protomers can be separated by preparative gel electrophoresis [9]. The crystal structure of circular N_m-RNA complexes from RV (m = 11) and VSV (m = 10) revealed that N is a two-domain protein that enwraps completely the viral RNA molecule and hides it from the polymerase and from the innate immune system [10, 11].

P is essential for viral replication and plays multiple roles at different stages of the viral cycle [12]. Firstly, P acts as a chaperone of N, by forming N^0-P complexes (where the superscript ^{0} denotes the absence of RNA) thus preserving N in a RNA-free form until N is transferred, by an unknown mechanism, from P to the nascent viral RNA in order to form new NCs [13-15]. Secondly, P forms with the L protein a two-subunit RNA-dependent RNA polymerase complex [12, 16]. L carries out all enzymatic activities of transcription and replication, namely RNA synthesis, mRNA cap synthesis and mRNA polyadenylation, while P acts as an essential cofactor although its precise role remains largely unknown. The only identified role of P in this complex is to mediate the attachment of the L subunit to the N-RNA template. L makes no direct interaction with N, but P binds to the N-RNA template through its C-terminal domain and to the L subunit through its N-terminal part [17], and thus provides a physical link between the polymerase and its N-RNA template. In order to synthesize both messenger and genomic RNA, the L polymerase must move along its N-RNA template. L is a processive enzyme [18] and P must maintain the attachment of L to its template during its displacements.

Recent studies have shown that P and N proteins from viruses of the *Paramyxoviridae* contain long Intrinsically Disordered Regions (IDRs) that alternate with structured domains [19-24]. Both amino acid sequence analysis and experimental data also suggested the presence of IDRs in P and N proteins from RV and VSV [19, 25-27]. The P protein from both viruses is highly sensitive to proteolysis and struc-

*Address correspondence to this author at the Unit of Virus Host Cell Interactions (UVHCI), UMI 3265 UJF-EMBL-CNRS, 6, rue Jules Horowitz, B.P. 181, 38042 Grenoble Cedex 9, France; Tel: + 33 4 76 20 94 62; Fax: + 33 4 76 20 94 00; E-mail: jamin@embl.fr

0929-8665/10 $55.00+.00

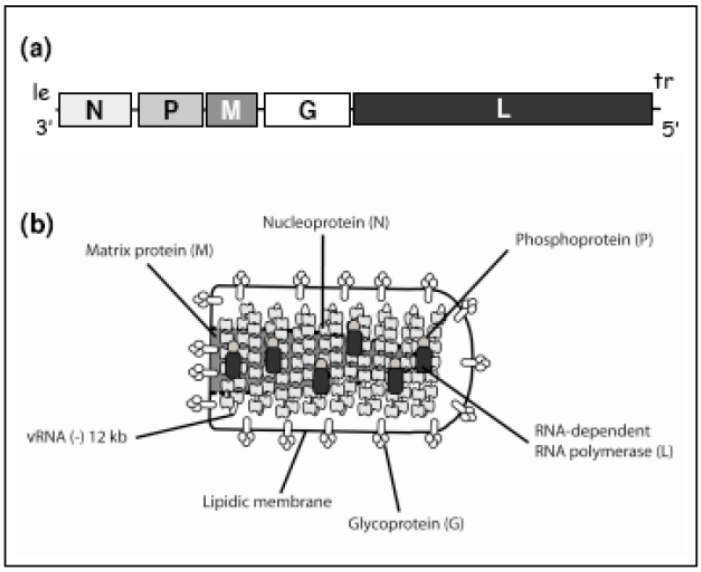


Figure 1. Genome organization and morphology of the *Rhabdoviridae*. (a) Genome organization. The gene products are denoted as follows: N, nucleoprotein, P, phosphoprotein, M, matrix protein, G, surface glycoprotein and L, large subunit of the RNA-dependent RNA polymerase complex. Le and Tr indicate the leader and trailer RNA sequences. (b) Schematic representation of a *Rhabdoviridae* particle. G is a glycoprotein embedded in the lipidic membrane, involved both in virus entry into host cells and in virus budding. M is the matrix protein, involved in assembly of viral particles and in virus budding. N, P and L plus the genomic RNA constitute the replication unit.

tured domains in RV and VSV P were identified by limited proteolysis [25-27]. In RV N, limited proteolysis eliminates the C-terminal part of the protein and abrogates its interaction with P, although it has no effect on the N-N interactions, also suggesting the presence of a disordered region in N [6, 28].

DISORDER PREDICTIONS USING THE D-SCORE

Recently, the structure/function paradigm was challenged with the discovery that many proteins are intrinsically disordered (IDPs) or contain intrinsically disordered regions (IDRs) under physiological conditions [29-32], and that disorder is required for, or participates in the biological function of proteins [29, 32-39]. Some proteins are disordered in isolation but fold into a well-defined structure upon binding to a partner [38, 40-43]. The recent appraisal of the importance of IDPs and IDRs in the protein world raised new issues such as understanding the role and influence of disorder in the stability, dynamics and interaction with molecular partners, and predicting the boundaries between disordered and structured regions from the amino acid sequence. To solve this latter problem, various algorithms were developed that rely on different approaches (for recent reviews, see [44, 45]). However, these different predictors yield different results, with large discrepancies in the number, position and length of the predicted disordered regions [46, 47]. These discrepancies argue for the existence of different types ("flavors") of disorder [45, 48], but raise the problem of choosing the best algorithm and/or strategy for locating the boundaries of disordered regions in a given protein [44, 45]. To avoid this choice, we recently devised a simple procedure that integrates predictions from multiple algorithms and relies on a simple binary voting procedure [17, 49]. The amino acid sequence is submitted to different algorithms available through WEB servers. For each prediction, each residue is ranked in a binary manner as ordered or disordered using the default thresholds set in the WEB servers and without accounting for the statistical weights provided by some servers [17]. A score of 0 is attributed to each residue predicted to be disordered, whereas a score of 1 is attributed to each one predicted to be in a structured region. For each residue, the values obtained from the different predictors are added and divided by the number of used algorithms in order to normalize the score. The calculated D-score (Figs. **2a, 3a** and **4c**) varies from 0 to 1, and consensus disordered regions are arbitrarily defined as regions with a normalized D-score $\leq$ 0.50. This approach was successful for locating known structured domains within the phosphoproteins of Sendai virus [21, 22, 25, 49, 50] and rabies virus [26] and for identifying unknown structured domains in VSV [49] and RV [51]. This analysis, backed by biochemical and biophysical data led to a model of modular organization of *Rhabdoviridae* P proteins [17, 52].

MODULAR ORGANIZATION OF *RHABDOVIRIDAE* P

On the basis of disorder predictions and of the structural characterization of different domains, a structural map is emerging. The D-score (Figs. **2a** and **3a**) calculated from the amino acid sequence suggests the presence of four structured regions separated by three IDRs in VSV P (Figs. **2b** and **2c**) and of three structured regions separated by two IDRs in RV

P (Figs. **3b** and **3c**), respectively. A comparison between VSV and RV P revealed the existence of common functional and structural regions, and suggested a modular organization for *Rhabdoviridae* P [17] similar to that previously proposed for *Paramyxoviridae* P [19].

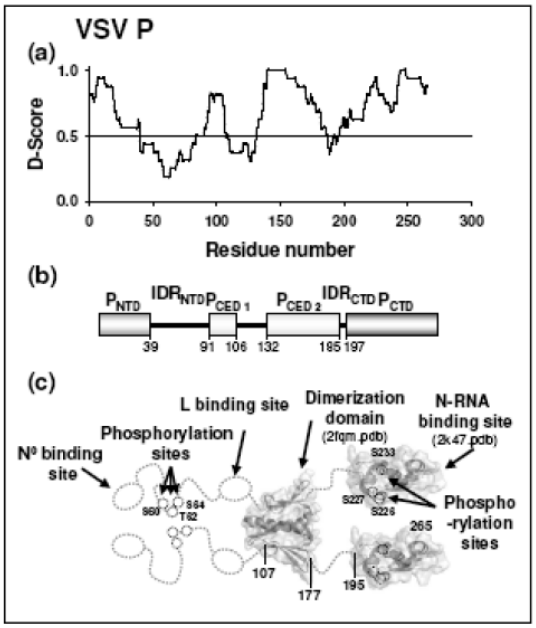


Figure 2. Modular organization of VSV P. (a) Score for disorder (D-score) as a function of residue number. Consensus disordered regions are defined as regions with a D-score < 0.50. (b) Location of consensus disordered regions with a D-score < 0.50. Numbers below the boxes indicate the boundaries of the consensus disordered regions. (c) Structural model showing the known three-dimensional structures as ribbon diagrams within space-filled model, the IDR as dotted lines and the regions that putatively fold upon binding to partners as dotted ovals. The small circles show the position of phosphorylation sites involved in transcriptional regulation.

In both RV and VSV P, the D-score predicts a structured N-terminal region (P_{NTD}) encompassing the first 40 to 60 residues. This region contains the binding site for N^0 as protein fragments corresponding to P_{NTD} of both VSV P and RV P are sufficient for maintaining N in a soluble, RNA-free form [14, 15]. However, in isolation these fragments are disordered [17]. The Stokes's radius of P_{NTD} measured by size exclusion chromatography is that of a fully unfolded protein of this size, and circular dichroism spectroscopy reveals low amounts of secondary structure [17]. This part of the protein is highly acidic and could mimic the RNA in its electrostatic interactions with N^0 and could fold only upon binding to N^0. The RNA binding groove in N is narrow (Fig. **4a**) and inserting P_{NTD} into this cavity might require it to be unfolded in the initial stages of the binding process.

The D-score predicts two structured domains in the central part of VSV P (P_{CED1}: amino acid 91-106 and P_{CED2}:

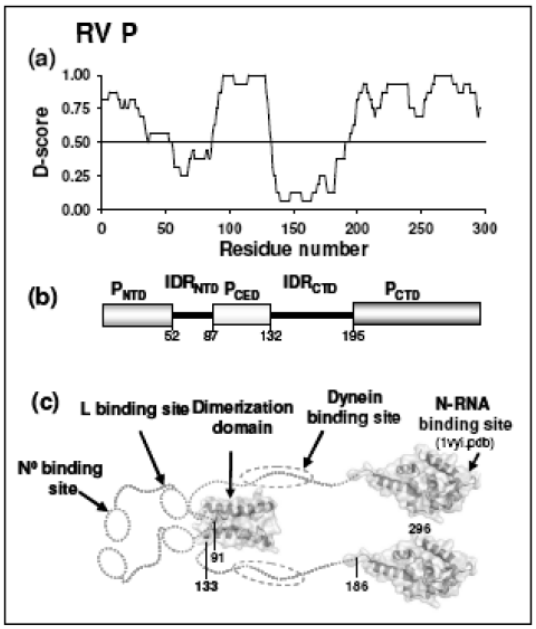


Figure 3. Modular organization of RV P. (a) Score for disorder (D-score) as a function of residue number. Consensus disordered regions are defined as regions with a D-score < 0.50. (b) Location of consensus disordered regions with a D-score < 0.50. Numbers below the boxes indicate the boundaries of the consensus disordered regions. (c) Structural model showing the known three-dimensional structures as ribbon diagrams within space-filled model, the IDR as dotted lines and the regions that putatively fold upon binding to partners as dotted ovals.

amino acid 132-185), and one structured region in that of RV P (P_{CED}: amino acid 87-132) (Fig. **3a** and **3b**). VSV P_{CED1} is conserved among VSV strains [17] and is part of one of the binding sites proposed for L [53]. There is no experimental evidence that P_{CED1} is folded in free P [27], and it may fold only upon binding to its partner. No equivalent structured or conserved region is found in RV P [17], although a major binding site for L has been similarly located in the 40-100 region by two-hybrid analysis [54]. In addition, peptides corresponding to the first 57 or 60 residues of RV P, and thus overlapping the N^0 and L binding sites, inhibit the transcription/replication of a rabies virus minigenome as well as rabies infection in kidney cells and in two neuronal cell lines [54]. This study however did not clearly demonstrate whether the inhibition is caused by inhibiting the formation of N^0-P or by preventing the binding to L. Both P proteins form oligomers [52]. For VSV, it has been proposed that phosphorylation of P controls transcription by modifying the oligomeric state of P [55, 56]. Size exclusion chromatography combined with detection by multi-angle laser light scattering (SEC-MALLS) and small-angle neutron scattering (SANS) revealed that recombinant P proteins form dimers in solution [52]. Previous studies showed that the region responsible for the oligomerization is localized in the central part of P (VSV P, amino acid 107-177; RV P, amino acid 52

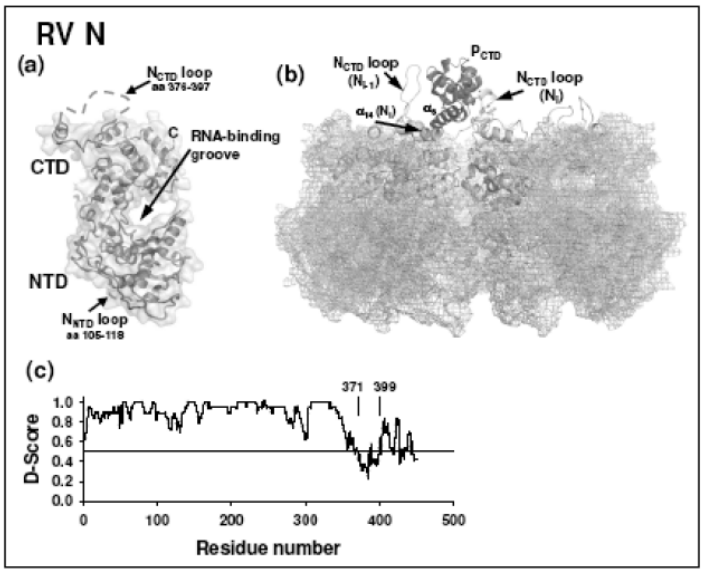


Figure 4. The Flexible N_{CTD} loop participates in binding P_{CTD}. (a) Ribbon diagram within a space-filled model of the N protomer in a lateral view showing that the RNA binding groove is closed. The N_{NTD} and N_{CTD} loops are shown as dotted lines. (b) Structure of the complex formed between RV P_{CTD} and the N_{11}–RNA complex obtained by molecular docking and modeling. The ribbon representation shows the location of P_{CTD} at the surface of the circular N_{11}–RNA ring. The meshed structure shows the circular N_{11}–RNA complex. The light gray ribbon representations show the three N_{CTD} domains used in the molecular modeling and the dark gray ribbon representation shows the P_{CTD}. The N_{CTD} loop of protomers N_i and N_{i-1} are shown, as well as helix α_6 of P_{CTD} and helix α_{14} of the N_i protomer. (c) D-score calculated for RV N.

to 189) [3, 57, 58]. The central region of VSV P corresponding to residues 107 to 177 was identified as a folded domain by limited proteolysis [27] and the crystal structure revealed that it forms dimers [25]. The monomer is made of a central α-helix preceded and followed by a two-stranded β-sheet. In the dimer, the helices from each monomer pack against each other in a parallel orientation, and at both sides of the helices, the N-terminal β-sheet of one monomer forms a four stranded β-sheet with the C-terminal β-sheet of the other monomer [52] (Fig. **2c**). The first part of the dimerization domain, forming the N-terminal β-sheet, is predicted to be disordered by the D-score (amino acid 107-131). However, in the crystal structure, it appears to be folded, stabilized by intermolecular interactions, suggesting that it could fold only upon forming the dimer. The first β-strand is rich in charged residues and poor in hydrophobic residues, sequence features that are typically associated with disorder, and which could explain why most disorder predictors, and consequently the D-score, stated this part as disordered. In RV P, the D-score locates more precisely the central domain within residues 87 to 132. Two-hybrid experiments confirmed that this region contains the dimerization domain [17]. The isolated central region of RV P forms an autonomous folding unit and SEC-MALLS experiments confirmed that it forms dimers in solution, whereas the protein deleted from this central region ($P_{\Delta 91\text{-}131}$) forms monomers [17]. Recently, the structure of the dimerization domain (amino acid 91-133) was solved by X-ray crystallography [51]. The monomer is made of two α-helices forming a hairpin, and the dimer is assembled by packing two monomers in a parallel orientation with an angle of about 50°. This structure is so different from that of VSV P_{CED2} that it is impossible to determine whether these domains have evolved from a common ancestor [51]. The largest functional difference between the two dimerization domains is that in the RV P dimer the N-terminal end is positioned next to the C-terminal end whereas in the VSV P dimer the N- and C-terminal ends are on opposite sides of the dimerization domain. However, this difference may not have biological consequences because the N- and C-terminal domains are connected to the dimerization domain by disordered regions [17], allowing a similar topological arrangement of the domains (Figs. **2c** and **3c**).

The last predicted structured region in RV P (amino acid 195-297) corresponds closely to the C-terminal N-RNA binding domain (P_{CTD}), whose structure was solved by X-ray crystallography (amino acid 186-296) [26] (Fig. **3c**). The equivalent domain was correctly localized in VSV P from the D-score analysis (amino acid 197-265), the isolated domain (amino acid 195-265) behaved as an autonomous folding unit and its structure was solved by NMR [49]. A longer form of this domain (amino acid 185-265) showed that the preceding ten residues (amino acid 185-194) are disordered, confirming the existence of the IDR_{CTD} [49]. Although multiple sequence alignments indicate almost no sequence similarity between P proteins from VSV and RV, the structure of VSV P_{CTD} is homologous to that of RV P. The topology of the backbone is conserved, although VSV P_{CTD} lacks helices

α_3 and α_6 of RV P_{CTD} [49]. This structure suggested that the C-terminal domain and its associated biological functions are conserved among the *Rhabdoviridae*, a proposal recently confirmed by the structure of mokola virus P_{CTD} [59].

FLEXIBLE LOOPS IN N PARTICIPATE IN BINDING OF P

In the crystal structure of circular N-RNA complexes from RV and VSV, the N protomer consists of two main domains, which completely enwrap the single-stranded RNA [10, 11] (Fig. **4a**). The N-terminal core domain (NTD) folds into a helical arrangement composed of helices connected by large loops. The C-terminal core domain (CTD) is composed of helices joined by tighter loops. Both the NTD and CTD act as jaws that clamp down onto the RNA strand and enclose it completely, an observation which is consistent with the fact that the RNA remains bound to N in CsCl gradients [60]. In addition to the core domains, two smaller subdomains participate in domain exchange between protomers and stabilize polymeric forms [10, 11].

In the RV structure, two regions, the N_{NTD} loop (amino acid 105 to 118) and the N_{CTD} loop (amino acid: 376 to 397), are flexible and absent in the crystal structure. Limited trypsin digestion cleaves the protein at Lys376 [28]. Trypsinated N-RNA complexes lose their ability to bind P_{CTD} although they retain their overall structure [6, 28]. The calculated D-score predicts a disordered region between residue 371 and residue 399 (Fig. **4c**), in good agreement with the location of the flexible N_{CTD} loop.

Recently, the structure of the complex between the N-RNA template and the C-terminal domain of P were obtained for both RV and VSV. A multiple stage flexible docking procedure was used to locate the binding site for the C-terminal domain of P on the N-RNA complex of RV and to predict the structure of the complex [61] in agreement with biological data [57]. The structure of the equivalent N-RNA-PCTD complex of VSV was determined by X-crystallography [62]. In both VSV and RV complexes, the P_{CTD} lies on the top of a N protomer (N_i) with the C-terminal helix of PCTD (α_6 in RV, α_4 in VSV) lying on the antepenultimate helix (α14 in RV, α_{13} in VSV) of N_i [10, 11, 63]. By an induced-fit mechanism, the NCTD loops of the same protomer (N_i) and of the adjacent one (N_{i-1}) mold around the PCTD, making extensive protein-protein contacts that could explain the submicromolar affinity of P for its template [61]. In both VSV and RV, the P_{CTD} exhibits the same orientation on the top side of the N-RNA ring, its N-terminal extremity pointing toward the centre of the ring. For the RV complex, titration experiments revealed that binding sites for isolated P_{CTD} on the circular N_m-RNA complexes are independent.

The main differences between the RV and VSV complexes reside in the numbers of residues involved in the interaction and in the surface area buried in the complex. In the RV complex, 34 residues of P_{CTD} and 39 residues of N (28 in N_i and 11 in N_{i+1}) participate in the binding against 18 and 17 residues in the VSV P_{CTD} and N, respectively. The N_{CTD} loop is longer in RV than in VSV, and the loops from two adjacent N protomers form a larger clamp in RV that hides a larger part of P_{CTD} than in VSV. In the RV complex, the total surface area of P_{CTD} (total surface area = 6,465 Å^2) buried upon interaction with two adjacent N molecules is 1,739 Å^2 (~27% of the total surface area), significantly larger than in the VSV complex for which only 956 Å^2 are buried (~19% of the total surface area). The other important difference is in the occupancy of the potential P_{CTD} binding sites on the N-RNA ring. In the VSV crystal structure, one P_{CTD} is bound to each N protomer, although reduced occupancy was observed in some positions [62]. In the RV complex, the involvement of N_{CTD} loops from adjacent protomers in binding one P_{CTD} makes unlikely to reach a stoichiometry larger than one P_{CTD} for two N protomers.

Folding of an IDR upon binding to its partner is also involved in the attachment of P to the N-RNA template in *Paramyxoviridae* [64-66]. In this other *MNV* family, the C-terminal part of N, named N_{TAIL}, is an IDR containing a short region that transiently populates α-helical conformers [65] and that folds upon binding to the C-terminal P_X domain [21] (also, see reviews by Longhi and Oglesbee and by Jensen *et al.* in this issue).

ROLE OF STRUCTURAL DISORDER IN PROTEINS OF THE *RHABDOVIRIDAE* REPLICATION COMPLEX

Disordered regions have been found in many functional proteins [30] and have been associated with specific functions [32, 38]. Some of these roles can be assigned to disordered regions found in P and N proteins of both RV and VSV. Firstly, disordered regions have major roles in exposing sites of posttranslational modifications, such as phosphorylation [32, 37]. In VSV, phosphorylation of residues located in the IDR_{NTD} regulates transcriptional activation [55, 56]. Other phosphorylation sites located in the structured P_{CTD} have no known functional roles [67] (Fig. **2c**). Phosphorylation sites are also found in the IDR_{NTD} of RV P, although no role could be assigned to these modifications [58, 68]. Secondly, IDRs and IDPs have been shown to participate in multi-molecular assemblies by simply bridging together different partners, by regulating the spacing between binding sites, by providing a large binding surface or multiple contact points for a large protein or by adopting different conformations in response to different stimuli [32, 38]. Often IDRs fold upon binding to their partners, a process accompanied by a large decrease in conformational entropy that can provide a high specificity associated with a moderate binding strength [38, 69]. RV and VSV P play key roles in the assembly of the replication/transcription complex by binding to the L-polymerase, [53, 54, 70], to N in the N^0–P complex and to N in the N–RNA template [14, 15, 71-73]. The attachment of P to the N-RNA template involves the closure around P_{CTD} of the N_{CTD} loops from two adjacent N protomers [61, 62]. This peculiar binding site provides an explanation to why binding of P to N through P_{CTD} can only occur with the N-RNA complex. The IDR_{NTD} of RV and VSV P contains two binding sites for viral partners, one for N_0 and the other for the L polymerase. The binding site for N_0 (P_{NTD}) in RV and VSV and the binding site for L in VSV are predicted to be structured but appear as disordered in the absence of their partners, while the binding site for L in RV is located in the IDR_{NTD} and was thus not predicted as a structured region [17]. The involvement of IDRs in the interactions with these viral partners suggests that these re-

gions could fold upon binding. Because these interactions are transient during the replication process, folding upon binding could provide the high specificity required for the correct assembly of the viral machineries and a modest or moderate affinity allowing the rearrangement of these multimolecular complexes [38, 69]. In the case of RV P, the IDR_{CTD} also contains a binding site for the cellular dynein light chain LC8 [74, 75]. Finally, IDPs and IDRs can provide a mean for establishing long-range interactions between partner proteins or for increasing the speed of interaction through the so-called "fly-casting" mechanism [41]. The location of binding sites for N^0 and L in the long disordered N-terminal region (P_{NTD} + IDR_{NTD}) could serve for catching these viral components and for bringing them to the site of RNA synthesis on the N-RNA template.

Recent data led us to propose a new model for the *Rhabdoviridae* replication complex [17, 49, 52, 61]. The affinity of P_{CTD} for the N_m–RNA template is high, with dissociation constants in the submicromolar range (160 ± 20 nM) [61]. Moreover, the association and dissociation kinetics measured by surface plasmon resonance (SPR) are slow (tens of seconds time range) as compared to the time scale of viral replication [61]. Both observations are difficult to conciliate with a mechanism involving continuous association and dissociation of P_{CTD} from the N_m–RNA template. In addition, in *Rhabdoviridae* the central dimerization seems to be dispensable for RNA synthesis [57, 76]. Recently, we showed that RV P deleted from this central domain, $P_{\Delta 91-131}$, is monomeric [17] and this mutant is functional in a minigenome transcription assay (Tordo N., personal communication). These results suggest that P dimers are continuously attached to the N-RNA template in the virion and in the cytoplasm of infected cells (Fig. **5**). The model obtained for the RV N-RNA-P_{CTD} complex indicates that two protomers of N are required for binding one P_{CTD}, suggesting that one P dimer could attach every four N protomers, whereas the estimated number of P dimers in the VSV virion [77] would result in one P dimer attached every five N protomers. When P is attached to the N-RNA template through its C-terminal domains, the binding sites for N^0 and L, located in the N-terminal part of P could thus extend far from the N-RNA template to catch their partners (Fig. **5**). This could allow the polymerase to move along the N-RNA template by jumping between adjacent P dimers and also to bring N^0 to the site of assembly of the new NC. This mechanism contrast with the cartwheeling mechanism proposed for the *Paramyxoviridae* [1, 78]. In this model, P, which forms tetramers through its central coiled-coil domain, would rotate around its central coiled-coil axle with the four C-terminal PX domains continuously associating and dissociating from the N–RNA template and thus carrying along their L cargo. There are two main differences in this part of the replication machinery between *Rhabdoviridae* and *Paramyxoviridae*. Firstly, the oligomerization domain of P seems to be dispensable for the synthesis of RNA in *Rhabdoviridae*, whereas a mutant lacking the central tetramerization domain in Sendai virus is unable to synthesize RNA [1, 78]. Secondly, affinities of PX for N_{TAIL} measured by NMR titration or isothermal titration calorimetry yielded dissociation constants in the range of 10 to 50 µM for both measles virus and Sendai virus proteins [21, 64, 79-82] and are thus weaker that that found for RV. However, for measles virus values near 100 nM were also obtained by SPR [83] suggesting a more complex picture (see the review by Longhi and Oglesbee in this issue) .

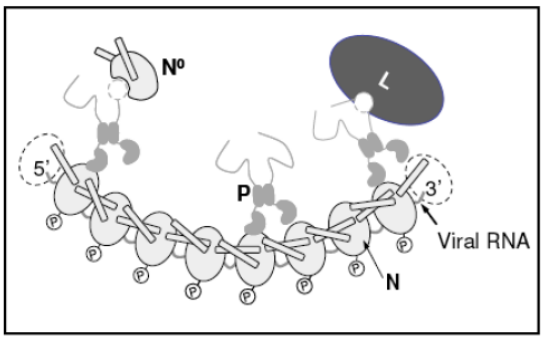


Figure 5. Schematic model for the attachment of P, L and N^0 to the N-RNA template. The modular P protein is attached at regular intervals along the N-RNA template through its C-terminal N-RNA binding domain. The N-terminal IDRs that contain the binding sites for both N^0 and L, are thus stretching away from the template, "fly-fishing" their partner proteins and bringing them to the template. Binding to their partners could be coupled to folding of the interacting regions (light gray circles). The bars represent the sub-domains of N interacting with the adjacent protomer on each side. The circled P indicate the putative phosphorylations of Ser389.

ACKNOWLEDGMENTS

This work was supported by the interdisciplinary program "Maladies Infectieuses Emergentes" from the CNRS and by a grant from the French ANR (ANR-07-001-01 (ANRAGE)) and Lyonbiopôle. The authors thank Danielle Blondel, Noël Tordo, Martin Blackledge, Malene Jensen, Thibaut Crépin, Giuseppe Zaccai and Sonia Longhi for fruitful discussions and collaborations, and the Partnership for Structural Biology for the excellent structural biology environment.

ABBREVIATIONS

MNV	=	*Mononegavirales* order
VSV	=	Vesicular stomatitis virus
RV	=	Rabies virus
RNA	=	Ribonucleic acid
N	=	Nucleoprotein
N-RNA	=	Nucleoprotein-ribonucleic acid complex
P	=	Phosphoprotein
L	=	Large subunit of the RNA-dependent RNA polymerase
NC	=	Nucleocapsid
N^0-P	=	Complex between the RNA-free nucleoprotein and the phosphoprotein
P_{CTD}	=	C-terminal domain of P
P_{CED}	=	Central domain of P

P_{NTD} = N-terminal domain of P

N_{NTD} = N-terminal domain of N

N_{CTD} = C-terminal domain of N

IDR = Intrinsically disordered region

IDP = Intrinsically disordered protein

SEC = Size exclusion chromatography

MALLS = Multi-angle laser light scattering

SANS = Small-angle neutron scattering

SAXS = Small-angle X-ray scattering

NMR = Nuclear magnetic resonance

SPR = Surface plasmon resonance.

REFERENCES

[1] Kolakofsky, D.; Le Mercier, P.; Iseni, F.; Garcin, D. Viral DNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology*, **2004**, *318*, 463-473.

[2] Whelan, S. P.; Barr, J. N.; Wertz, G. W. Transcription and replication of nonsegmented negative-strand RNA viruses. *Curr. Top. Microbiol. Immunol.*, **2004**, *283*, 61-119.

[3] Pattnaik, A. K.; Hwang, L.; Li, T.; Englund, N.; Mathur, M.; Das, T.; Banerjee, A. K. Phosphorylation within the amino-terminal acidic domain I of the phosphoprotein of vesicular stomatitis virus is required for transcription but not for replication. *J. Virol.*, **1997**, *71*, 8167-8175.

[4] Gupta, A. K.; Shaji, D.; Banerjee, A. K. Identification of a novel tripartite complex involved in replication of vesicular stomatitis virus genome RNA. *J. Virol.*, **2003**, *77*, 732-738.

[5] Arnheiter, H.; Davis, N. L.; Wertz, G.; Schubert, M.; Lazzarini, R. A. Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. *Cell*, **1985**, *41*, 259-267.

[6] Schoehn, G.; Iseni, F.; Mavrakis, M.; Blondel, D.; Ruigrok, R. W. Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site. *J. Virol.*, **2001**, *75*, 490-498.

[7] Green, T. J.; Macpherson, S.; Qiu, S.; Lebowitz, J.; Wertz, G. W.; Luo, M. Study of the assembly of vesicular stomatitis virus N protein: role of the P protein. *J. Virol.*, **2000**, *74*, 9515-9524.

[8] Iseni, F.; Barge, A.; Baudin, F.; Blondel, D.; Ruigrok, R. W. Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures. *J. Gen. Virol.*, **1998**, *79 (Pt 12)*, 2909-2919.

[9] Albertini, A. A.; Clapier, C. R.; Wernimont, A. K.; Schoehn, G.; Weissenhorn, W.; Ruigrok, R. W. Isolation and crystallization of a unique size category of recombinant Rabies virus Nucleoprotein-RNA rings. *J. Struct. Biol.*, **2006**, *158*, 129-133.

[10] Albertini, A. A.; Wernimont, A. K.; Muziol, T.; Ravelli, R. B.; Clapier, C. R.; Schoehn, G.; Weissenhorn, W.; Ruigrok, R. W. Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science*, **2006**, *313*, 360-363.

[11] Green, T. J.; Zhang, X.; Wertz, G. W.; Luo, M. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science*, **2006**, *313*, 357-360.

[12] Emerson, S. U.; Yu, Y. Both NS and L proteins are required for *in vitro* RNA synthesis by vesicular stomatitis virus. *J Virol*, **1975**, *15*, 1348-1356.

[13] Masters, P. S.; Banerjee, A. K. Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA. *J. Virol.*, **1988**, *62*, 2658-2664.

[14] Mavrakis, M.; Mehouas, S.; Real, E.; Iseni, F.; Blondel, D.; Tordo, N.; Ruigrok, R. W. Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. *Virology*, **2006**, *349*, 422-429.

[15] Chen, M.; Ogino, T.; Banerjee, A. K. Interaction of vesicular stomatitis virus P and N proteins: Identification of two overlapping domains at the N-terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA. *J. Virol.*, **2007**, *81*, 13478-13485.

[16] De, B. P.; Banerjee, A. K. Specific interactions of vesicular stomatitis virus L and NS proteins with heterologous genome ribonucleoprotein template lead to mRNA synthesis *in vitro*. *J. Virol.*, **1984**, *51*, 628-634.

[17] Gérard, F. C. A.; Ribeiro , E. A.; Leyrat, C.; Ivanov, I.; Blondel, D.; Longhi, S.; Ruigrok, W. H.; Jamin, M. Modular organization of rabies virus phosphoprotein. *J. Mol. Biol.*, **2009**, *388*, 978-996.

[18] Stillman, E. A., and Whitt, M. Transcript initiation and 5'-end modifications are separable events during vesicular stomatitis virus transcription. *J. Virol.*, **1999**, *73*, 7199-7209.

[19] Karlin, D.; Ferron, F.; Canard, B.; Longhi, S. Structural disorder and modular organization in Paramyxovirinae N and P. *J. Gen. Virol.*, **2003**, *84*, 3239-3252.

[20] Longhi, S. Nucleocapsid structure and function. *Curr. Top. Microbiol. Immunol.*, **2009**, *329*, 103-128.

[21] Blanchard, L.; Tarbouriech, N.; Blackledge, M.; Timmins, P.; Burmeister, W. P.; Ruigrok, R. W.; Marion, D. Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. *Virology*, **2004**, *319*, 201-211.

[22] Tarbouriech, N.; Curran, J.; Ruigrok, R. W.; Burmeister, W. P. Tetrameric coiled coil domain of Sendai virus phosphoprotein. *Nat. Struct. Biol.*, **2000**, *7*, 777-781.

[23] Llorente, M. T.; Garcia-Barreno, B.; Calero, M.; Camafeita, E.; Lopez, J. A.; Longhi, S.; Ferron, F.; Varela, P. F.; Melero, J. A. Structural analysis of the human respiratory syncytial virus phosphoprotein: characterization of an alpha-helical domain involved in oligomerization. *J. Gen. Virol.*, **2006**, *87*, 159-169.

[24] Bourhis, J. M.; Canard, B.; Longhi, S. Structural disorder within the replicative complex of measles virus: functional implications. *Virology*, **2006**, *344*, 94-110.

[25] Ding, H.; Green, T. J.; Lu, S.; Luo, M. Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J. Virol.*, **2006**, *80*, 2808-2814.

[26] Mavrakis, M.; McCarthy, A. A.; Roche, S.; Blondel, D.; Ruigrok, R. W. Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. *J. Mol. Biol.*, **2004**, *343*, 819-831.

[27] Ding, H., Green, T. J., and Luo, M. Crystallization and preliminary X-ray analysis of a proteinase-K-resistant domain within the phosphoprotein of vesicular stomatitis virus (Indiana). *Acta Crystallogr D Biol Crystallogr*, **2004**, *60*, 2087-2090.

[28] Kouznetzoff, A.; Buckle, M.; Tordo, N. Identification of a region of the rabies virus N protein involved in direct binding to the viral RNA. *J. Gen. Virol.*, **1998**, *79 (Pt 5)*, 1005-1013.

[29] Wright, P. E.; Dyson, H. J. Intrinsically unstructured proteins: reassessing the protein structure-function paradigm. *J. Mol. Biol.*, **1999**, *293*, 321-331.

[30] Uversky, V. N.; Gillespie, J. R.; Fink, A. L. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*, **2000**, *41*, 415-427.

[31] Uversky, V. N. What does it mean to be natively unfolded? *Eur. J. Biochem.*, **2002**, *269*, 2-12.

[32] Tompa, P. Intrinsically unstructured proteins. *Trends Biochem Sci*, **2002**, *27*, 527-533.

[33] Liu, J.; Perumal, N. B.; Oldfield, C. J.; Su, E. W.; Uversky, V. N.; Dunker, A. K. Intrinsic disorder in transcription factors. *Biochemistry*, **2006**, *45*, 6873-6888.

[34] Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z. Intrinsic disorder and protein function. *Biochemistry*, **2002**, *41*, 6573-6582.

[35] Dunker, A. K.; Brown, C. J.; Obradovic, Z. Identification and functions of usefully disordered proteins. *Adv. Protein Chem.*, **2002**, *62*, 25-49.

[36] Romero, P.; Obradovic, Z.; Dunker, A. K. Natively disordered proteins: functions and predictions. *Appl. Bioinformatics*, **2004**, *3*, 105-113.

[37] Iakoucheva, L. M.; Radivojac, P.; Brown, C. J.; O'Connor, T. R.; Sikes, J. G.; Obradovic, Z.; Dunker, A. K. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.*, **2004**, *32*, 1037-1049.

[38] Dyson, H. J.; Wright, P. E. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.*, **2002**, *12*, 54-60.

[39] Tompa, P., and Csermely, P. The role of structural disorder in the function of RNA and protein chaperones. *Faseb J.*, **2004**, *18*, 1169-1175.

[40] Spolar, R. S., and Record, M. T., Jr. Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **1994**, *263*, 777-784.
[41] Sugase, K.; Dyson, H. J.; Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, **2007**, *447*, 920-921.
[42] Demarest, S. J.; Martinez-Yamout, M.; Chung, J.; Chen, H.; Xu, W.; Dyson, H. J.; Evans, R. M., and Wright, P. E. Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature*, **2002**, *415*, 549-553.
[43] Uversky, V. N.; Segel, D. J.; Doniach, S.; Fink, A. L. Association-induced folding of globular proteins. *Proc Natl Acad Sci U S A*, **1998**, *95*, 5480-5483.
[44] Bourhis, J. M., Canard, B., and Longhi, S. Predicting protein disorder and induced folding: from theoretical principles to practical applications. *Curr. Protein Pept. Sci.*, **2007**, *8*, 135-149.
[45] Ferron, F., Longhi, S., Canard, B., and Karlin, D. A practical overview of protein disorder prediction methods. *Proteins*, **2006**, *65*, 1-14.
[46] Lieutaud, P.; Canard, B.; Longhi, S. MeDor: a metaserver for predicting protein disorder. *BMC Genomics*, **2008**, *9(Suppl. 2)*, S25.
[47] Ishida, T.; Kinoshita, K. Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics*, **2008**, *24*, 1344-1348.
[48] Vucetic, S.; Brown, C. J.; Dunker, A. K.; Obradovic, Z. Flavors of protein disorder. *Proteins*, **2003**, *52*, 573-584.
[49] Ribeiro, E. A.; Jr., Favier, A.; Gérard, F. C.; Leyrat, C.; Brutscher, B.; Blondel, D.; Ruigrok, R. W.; Blackledge, M.; Jamin, M. Solution Structure of the C-Terminal Nucleoprotein-RNA Binding Domain of the Vesicular Stomatitis Virus Phosphoprotein. *J. Mol. Biol.*, **2008**, *382*, 525-538.
[50] Bernado, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W.; Blackledge, M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc. Natl. Acad. Sci., U S A*, **2005**,
[51] Ivanov, I.; Crepin, T.; Jamin, M.; Ruigrok, R. Structure of the dimerisation domain of the rabies virus phosphoprotein. *J. Virol.* , **2010**, Accepted for publication.
[52] Gérard, F.; Ribeiro, E.; Albertini, A.; Zaccai, G.; Ebel, C.; Ruigrok, R.; Jamin, M. Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution. *Biochemistry*, **2007**, *46*, 10328-10338.
[53] Emerson, S. U.; Schubert, M. Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus. *Proc. Natl. Acad .Sci. U S A*, **1987**, *84*, 5655-5659.
[54] Castel, G.; Chteoui, M.; Caignard, G.; Prehaud, C.; Mehouas, S.; Real, E.; Jallet, C.; Jacob, Y.; Ruigrok, R. W.; Tordo, N. Peptides that mimic the amino-terminal end of the rabies virus phosphoprotein have antiviral activity. *J. Virol.*, **2009**, *83*, 10808-10820.
[55] Gao, Y.; Lenard, J. Multimerization and transcriptional activation of the phosphoprotein (P) of vesicular stomatitis virus by casein kinase-II. *Embo. J.*, **1995**, *14*, 1240-1247.
[56] Takacs, A. M.; Barik, S.; Das, T.; Banerjee, A. K. Phosphorylation of specific serine residues within the acidic domain of the phosphoprotein of vesicular stomatitis virus regulates transcription *in vitro*. *J. Virol.*, **1992**, *66*, 5842-5848.
[57] Jacob, Y.; Real, E.; Tordo, N. Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains. *J. Virol.*, **2001**, *75*, 9613-9622.
[58] Gigant, B.; Iseni, F.; Gaudin, Y.; Knossow, M.; Blondel, D. Neither phosphorylation nor the amino-terminal part of rabies virus phosphoprotein is required for its oligomerization. *J. Gen. Virol.*, **2000**, *81*, 1757-1761.
[59] Assenberg, R.; Delmas, O.; Ren, J.; Vidalain, P. O.; Verma, A.; Larrous, F.; Graham, S. C.; Tangy, F.; Grimes, J. M.; Bourhy, H. The Structure of the N-RNA Binding Domain of the Mokola virus Phosphoprotein. *J. Virol.*, **2009**, Nov 11. [Epub ahead of print]PMID: 19906936 [PubMed - as supplied by publisher].
[60] Lynch, S.; Kolakofsky, D. Ends of the RNA within Sendai virus defective interfering nucleocapsids are not free. *J. Virol.*, **1978**, *28*, 584-589.
[61] Ribeiro, E. D.; Jr., Leyrat, C.; Gerard, F. C.; Albertini, A. A.; Falk, C.; Ruigrok, R. W.; Jamin, M. Binding of Rabies Virus Polymerase Cofactor to Recombinant Circular Nucleoprotein-RNA Complexes. *J. Mol. Biol.*, **2009**, *394*, 558-575.
[62] Green, T. J.; Luo, M. Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P. *Proc. Natl. Acad. Sci. U S A*, **2009**, *106*, 11721-11726.
[63] Albertini, A. A., Schoehn, G., Weissenhorn, W., and Ruigrok, R. W. Structural aspects of rabies virus replication. *Cell Mol Life Sci*, **2008**, *65*, 282-294.
[64] Kingston, R. L.; Hamel, D. J.; Gay, L. S.; Dahlquist, F. W.; Matthews, B. W. Structural basis for the attachment of a paramyxoviral polymerase to its template. *Proc. Natl. Acad. Sci. U S A*, **2004**, *101*, 8301-8306.
[65] Jensen, M. R.; Houben, K.; Lescop, E.; Blanchard, L.; Ruigrok, R. W.; Blackledge, M. Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. *J. Am. Chem. Soc.*, **2008**, *130*, 8055-8061.
[66] Longhi, S.; Receveur-Brechot, V.; Karlin, D.; Johansson, K.; Darbon, H.; Bhella, D.; Yeo, R.; Finet, S.; Canard, B. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *J. Biol. Chem.*, **2003**, *278*, 18638-18648.
[67] Gao, Y.; Lenard, J. Cooperative binding of multimeric phosphoprotein (P) of vesicular stomatitis virus to polymerase (L) and template: pathways of assembly. *J. Virol.*, **1995**, *69*, 7718-7723.
[68] Gupta, A. K.; Blondel, D.; Choudhary, S.; Banerjee, A. K. The phosphoprotein of rabies virus is phosphorylated by a unique cellular protein kinase and specific isomers of protein kinase C. *J. Virol.*, **2000**, *74*, 91-98.
[69] Wright, P. E.; Dyson, H. J. Linking folding and binding. *Curr. Opin. Struct. Biol.*, **2009**, *19*, 31-38.
[70] Chenik, M.; Schnell, M.; Conzelmann, K. K.; Blondel, D. Mapping the interacting domains between the rabies virus polymerase and phosphoprotein. *J. Virol.*, **1998**, *72*, 1925-1930.
[71] Chenik, M.; Chebli, K.; Gaudin, Y.; Blondel, D. *In vivo* interaction of rabies virus phosphoprotein (P) and nucleoprotein (N): existence of two N-binding sites on P protein. *J. Gen. Virol.*, **1994**, *75* 2889-2896.
[72] Masters, P. S.; Banerjee, A. K. Resolution of multiple complexes of phosphoprotein NS with nucleocapsid protein N of vesicular stomatitis virus. *J. Virol.*, **1988**, *62*, 2651-2657.
[73] Takacs, A. M.; Das, T.; Banerjee, A. K. Mapping of interacting domains between the nucleocapsid protein and the phosphoprotein of vesicular stomatitis virus by using a two-hybrid system. *Proc. Natl. Acad. Sci. U S A*, **1993**, *90*, 10375-10379.
[74] Raux, H.; Flamand, A.; Blondel, D. Interaction of the rabies virus P protein with the LC8 dynein light chain. *J. Virol.*, **2000**, *74*, 10212-10216.
[75] Jacob, Y.; Badrane, H.; Ceccaldi, P. E.; Tordo, N. Cytoplasmic dynein LC8 interacts with lyssavirus phosphoprotein. *J. Virol.*, **2000**, *74*, 10217-10222.
[76] Das, S. C., and Pattnaik, A. K. Role of the hypervariable hinge region of phosphoprotein P of vesicular stomatitis virus in viral RNA synthesis and assembly of infectious virus particles. *J. Virol.*, **2005**, *79*, 8101-8112.
[77] Thomas, D.; Newcomb, W. W.; Brown, J. C.; Wall, J. S.; Hainfeld, J. F.; Trus, B. L.; Steven, A. C. Mass and molecular composition of vesicular stomatitis virus: a scanning transmission electron microscopy analysis. *J. Virol.*, **1985**, *54*, 598-607.
[78] Curran, J.; Kolakofsky, D. Replication of paramyxoviruses. *Adv. Virus Res.*, **1999**, *54*, 403-422.
[79] Kingston, R. L.; Baase, W. A.; Gay, L. S. Characterization of nucleocapsid binding by the measles virus and mumps virus phosphoproteins. *J. Virol.*, **2004**, *78*, 8630-8640.
[80] Bernard, C.; Gely, S.; Bourhis, J. M.; Morelli, X.; Longhi, S.; Darbon, H. Interaction between the C-terminal domains of N and P proteins of measles virus investigated by NMR. *FEBS Lett.*, **2009**, *583*, 1084-1089.
[81] Houben, K.; Blanchard, L.; Blackledge, M.; and Marion, D. Intrinsic dynamics of the partly unstructured PX domain from the Sendai virus RNA polymerase co-factor P. *Biophys. J.*, **2007**, *93*, 2930-2940.
[82] Houben, K.; Marion, D.; Tarbouriech, N.; Ruigrok, R. W.; Blanchard, L. Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. *J. Virol.*,, **2007**, *81*, 6807-6816.

[83] Bourhis, J. M.; Receveur-Brechot, V.; Oglesbee, M.; Zhang, X.; Buccellato, M.; Darbon, H.; Canard, B.; Finet, S.; Longhi, S. The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded. *Protein Sci.*, **2005**, *14*, 1975-1992.

Received: ?????????????? **Revised:** ?????????????? **Accepted:** ??????????????

II. ARTICLE VIII: Structure and plasticity of the peptidyl-prolyl isomerase Par27 of Bordetella pertussis revealed by X-ray diffraction and small-angle X-ray scattering

Auteurs: Bernard Clantin, Cédric Leyrat, Alex Wohlkönig, Hélène Hodak, Euripedes de Almeida Ribeiro Jr, Nicolas Martinez, Catherine Baud, Caroline Smet-Nocca, Vincent Villeret, Françoise Jacob-Dubuisson et Marc Jamin.

Publié en 2010 dans *Journal of Structural Biology*, Volume 169, Pages 253-265

Journal of Structural Biology 169 (2010) 253–265

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Structure and plasticity of the peptidyl-prolyl isomerase Par27 of *Bordetella pertussis* revealed by X-ray diffraction and small-angle X-ray scattering

Bernard Clantin [c], Cédric Leyrat [a], Alex Wohlkönig [c], Hélène Hodak [b], Euripedes de Almeida Ribeiro Jr. [a], Nicolas Martinez [a], Catherine Baud [b], Caroline Smet-Nocca [d], Vincent Villeret [c], Françoise Jacob-Dubuisson [b,*], Marc Jamin [a,*]

[a] Unit of Virus Host Cell Interactions, UMI 3265 UJF-EMBL-CNRS, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France
[b] INSERM U629, IFR 142, Institut Pasteur de Lille, 1 rue Professeur Calmette, F-59019 Lille Cedex, France
[c] UMR8161 CNRS, Institut de Biologie de Lille – Institut Pasteur de Lille – Université de Lille 1 – Université de Lille 2, 1 rue Calmette, 59021 Lille Cedex, France
[d] CNRS UMR 8576, UGSF-IFR 147, Université des Sciences et Technologies de Lille 1, 59655 Villeneuve d'Ascq cedex, France

ARTICLE INFO

Article history:
Received 2 October 2008
Received in revised form 8 September 2009
Accepted 16 November 2009
Available online 20 November 2009

Keywords:
Chaperone
Peptidyl-prolyl isomerase
Bacterial periplasm
Whooping cough
Bordetella pertussis

ABSTRACT

Par27 from *Bordetella pertussis* belongs to a newly discovered class of dimeric peptidyl-prolyl isomerase (PPIase)/chaperones from the parvulin family. It is a tripartite protein with a central PPIase domain surrounded by N- and C-terminal sub-domains (NTD and CTD). Here, the Par27 structure was characterized by X-ray crystallography, small-angle X-ray scattering and template-based modeling. In the crystal lattice, Par27 consists of alternating well ordered and poorly ordered domains. The PPIase domains gave rise to diffuse scattering and could not be solved, whereas a 2.2 Å resolution crystal structure was obtained for the NTD and CTD, revealing a cradle-shaped dimeric platform. Despite a lack of sequence similarity with corresponding sub-domains, the topology of the peptide chain in the NTD/CTD core is similar to that of other monomeric PPIase/chaperones such as SurA and trigger factor from *Escherichia coli*. In Par27, dimerization occurs by sub-domain swapping. Because of the strong amino acid sequence similarity to other parvulin domains, a model for the Par27 PPIase domain was built by template-based modeling and validated against small-angle X-ray scattering (SAXS) data. A model of the full-length dimeric Par27 structure was built by rigid-body modeling and filtering against SAXS data using the partial crystal structure of the NTD/CTD core and the template-based PPIase model. The flexibility of protein was accounted for by representing the structure as an ensemble of different conformations that collectively reproduce the scattering data. The refined models exhibit a cradle-like shape reminiscent of other PPIase/chaperones, and the variability in the orientation of the PPIase domains relative to the NTD/CTD core platform observed in the different models suggests inter-domain flexibility that could be important for the biological activity of this protein.

1. Introduction

The periplasm of Gram-negative bacteria is a site of intense trafficking of proteins destined to the outer membrane, the cell surface or the extracellular milieu (Wülfing and Plückthun, 1994). Following their export across the cytoplasmic membrane, many such proteins cross that compartment in extended, unfolded conformations. Depending on their final destination, some of them acquire their native form in the periplasm, while others remain fully or partially extended prior to their secretion across the outer membrane. Gram-negative bacteria have thus developed periplasmic systems that provide assistance for protein folding and protection from aggregation or proteolytic degradation (Wülfing and Plückthun, 1994; Missiakas et al., 1996; Miot and Betton, 2004). These systems include broad-spectrum as well as specific chaperones, and folding catalysts such as disulfide oxidases/isomerases and peptidyl-prolyl isomerases (PPIases).

A large group of bacterial folding catalysts exhibit both chaperone and PPIase activities. The peptidyl-prolyl isomerase activity is carried out by small globular domains of approximately 100 amino acids that form three distinct families based on their amino acid sequence: the parvulins, the FK506-binding proteins (FKBP) and the cyclophilins (Rahfeld et al., 1994; Göthel and Marahiel, 1999). Each family comprises single-domain PPIases as well as multi-domain proteins. In the latter category, the additional domains often act as chaperones, increasing the affinity and providing the PPIase domain with selectivity for protein substrates (Scholz et al., 1997a). While the importance of the PPIase activity of such multi-domain proteins

* Corresponding authors. Address: Unit of Virus Host Cell Interactions (UVHCI), UMI 3265 UJF-EMBL-CNRS, 6 rue Jules Horowitz, B.P. 181, 38042 Grenoble Cedex 9, France. Fax: +33 4 76 20 94 00 (M. Jamin).
E-mail addresses: francoise.jacob@ibl.fr (F. Jacob-Dubuisson), jamin@embl.fr (M. Jamin).

1047-8477/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2009.11.007

in vivo remains unclear, the chaperone activity of some of them plays an essential role *in vivo* (Behrens et al., 2001). Thus, the chaperone activity of SurA in the periplasm of *Escherichia coli* is essential for outer membrane protein biogenesis (Lazar and Kolter, 1996; Rouviere and Gross, 1996; Lazar et al., 1998; Sklar et al., 2007).

We recently reported the identification of the prototype of a new group of parvulins, Par27, from the whooping cough agent *Bordetella pertussis* (Hodak et al., 2008). Par27 was identified in a search for accessory factors involved in the secretion of filamentous hemagglutinin (FHA), the major attachment factor of that bacterium. FHA is a large protein secreted by the two-partner secretion (TPS) pathway (Jacob-Dubuisson et al., 2004). It is exported through the cytoplasmic membrane by the Sec machinery and thus transits through the periplasm in an extended conformation before its transport across the outer membrane by FhaC, a specific transporter of the Omp85/TpsB superfamily (Clantin et al., 2007). At the cell surface, FHA progressively folds into a long amphipathic β-helix (Kajava et al., 2001; Clantin et al., 2004). Because FHA exposes amphipathic, extended structures in the periplasm prior to secretion and subsequent folding, it is a good model to study the role of periplasmic chaperones and catalysts. Par27 was shown to bind to a non-native FHA fragment by affinity chromatography (Hodak et al., 2008). Par27 also displays affinity for other proteins rich in amphipathic β structure such as outer membrane porins, and therefore it has been proposed to serve as a general periplasmic chaperone in *B. pertussis* (Hodak et al., 2008).

Par27, which exhibits both chaperone and PPIase activities *in vitro*, is the first identified parvulin protein that forms dimers in solution (Hodak et al., 2008). The central PPIase domain of Par27 is flanked by N- and C-terminal extensions that are found in a number of putative PPIases present mostly in β proteobacteria (Hodak et al., 2008). In this work, we have characterized the structure of Par27 by X-ray crystallography, template-based modeling and small-angle scattering (SAS). We show that the PPIase domains are attached to a homodimeric cradle-shaped 'core' domain made by the assembly of the N- (NTD) and C-terminal domains (CTD) of each monomer. By analogy with other PPIases/chaperones, the structure suggests that the NTD/CTD core constitutes the chaperone domain of the protein.

2. Materials and methods

2.1. Expression and purification of recombinant Par27 and its PPIase domain

Production of the full-length protein was described earlier (Hodak et al., 2008). For the expression of the peptidyl-prolyl isomerase (PPIase) domain of Par27 (aa 107–202), the central portion of the gene encoding Par27 was amplified by PCR with the primers 5′-gatccggcaagatggaatacaagg-3′ and 5′-aagcttttaacgggtatcgtcgacctggat-3′, and the resulting amplicon was inserted into the BamHI and HindIII sites of pQE30. Par27 and the isolated PPIase domain were produced in *E. coli* as recombinant proteins with a C-terminal His-tag. The proteins were purified by nickel affinity chromatography followed by gel filtration as previously described (Hodak et al., 2008). The protein concentrations were measured by absorbance at 280 nm using $\varepsilon_{280\ nm} = 18{,}500\ M^{-1}\ cm^{-1}$ for full-length Par27 and $14{,}000\ M^{-1}\ cm^{-1}$ for the PPIase domain.

The homogeneity of each sample was routinely checked by size exclusion chromatography (SEC) coupled with detection by multi-angle laser light scattering (MALLS) and refractometry (RI) (Wyatt, 1998). SEC was performed with a Superdex S200 column equilibrated in 20 mM Tris–HCl, 150 mM NaCl at pH 7.5 and 20 °C. MALLS detection was performed with a DAWN-EOS detector (Wyatt Technology Corp., Santa Barbara, CA) and the differential refractive index measurement was performed with a RI2000 detector (Schambeck SFD). A d*n*/d*c* value of 0.185 ml g^{-1} was used. Combined detection by MALLS and RI allows the determination of the weight-averaged molecular mass (M_w) and its distribution across the elution peak that informs about the dispersity of the sample (Hodak et al., 2008).

Far-UV circular dichroism spectra were recorded at 20 °C with a JASCO J-810 CD spectropolarimeter equipped with a temperature controller (Peltier system).

2.2. X-ray crystallography

Full-length Par27 was crystallized as reported previously (Wohlkönig et al., 2008). A single SeMet-labeled protein crystal was used to collect diffraction data at beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Data collection and refinement statistics are summarized in Tables 1 and 2. Three wavelengths (λ_{peak}, λ_{remote}, $\lambda_{inflexion}$) were chosen around the selenium K-absorption edge (0.9795 Å). The diffraction data for each set were indexed, integrated and scaled using XDS (Kabsch, 1993).

Data were indexed in space group C222, with cell parameters $a = 54.6$ Å, $b = 214.1$ Å, $c = 57.8$ Å, assuming two molecules in the asymmetric unit, or in space group *P*2 with unit-cell parameters $a = 42.2$ Å, $b = 142.8$ Å, $c = 56.0$ Å and $\beta = 95.1°$ (Wohlkönig et al., 2008). The positions of 4 of the 5 expected Se atoms were determined with ShelxD (Uson and Sheldrick, 1999). The Se atoms parameters were refined and initial experimental phases were calculated at 2.2 Å by the MAD or MIR method using Sharp (Bricogne et al., 2003). The density modification procedure and the automatic model building were also carried out using Sharp (Bricogne et al., 2003). Data were also deconvoluted using the program DETWIN (Collaborative Computational Project, Number 4, 1994. "The CCP4 Suite: Programs for Protein Crystallography" 1994). Molecular replacement was carried out using AMoRe (Trapani and Navaza, 2008).

2.3. Homology modeling

Homology modeling was performed by submitting the amino acid sequence to the meta-threading-server LOMETS (Wu and

Table 1
Data collection statistics.

Data collection	Se pk1
Cell parameters (Å)	54.90 215.56 57.81
(°)	90.0 90.0 90.0
Space group	*C*222
Wavelength (Å)	0.97910
Resolution (Å)[a]	2.2 (2.3–2.2)
Completeness (%)[a]	99.1 (100.0)
Redundancy[a]	3.78 (3.81)
I/sI[a]	15.36 (4.90)
R_{mrgd-F} (%)[a]	8.7 (33.5)
Beamline	ESRF ID23-1
Refinement	
R_{work} (%)[b]	30.72
R_{free} (%)	32.87
R.m.s.d.	
Bond lengths (Å)	0.029
Bond angles (°)	2.125
Ramachandran statistics	
Most favored (%)	90.2
Additionally allowed (%)	8.9
Generously allowed (%)	0.8
Disallowed (%)	0

[a] Number in parentheses is the statistic for the highest resolution shell.
[b] *R*-factor = $\Sigma||F_o| - |F_c||/\Sigma|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively.

Table 2
Experimental and theoretical molecular mass (MM) and molecular dimensions of full-length Par27 and of its isolated PPIase domain.

Proteins	MM (kDa) (from MALLS)	MM (kDa) (from SAXS)	MM (Da) (from sequence)	R_g (nm) (from Guinier)	R_g (nm) (from $P(r)$)	D_{max} (nm)
			(2 × 28,433)			
Full-length Par27	60 ± 5	55 ± 3	56,866	3.9 ± 0.1	3.8 ± 0.1	11.5
PPIase domain	12.0 ± 0.5	12 ± 1	12,222	1.7 ± 0.1	1.7 ± 0.1	5.9

Zhang, 2007), an automated server that takes predictions from nine threading servers, selects ten consensus models based on structure similarity and generates models with MODELLER. The Par27 PPIase domain was ranked as an 'easy' target because of the good quality of template alignments and the accuracy of constraints. Of the 10 consensus models, the first six were constructed using the PPIase domain of Pin1 (Bayer et al., 2003) as a template, and the next four using the PPIase domain of PrsA (Tossavainen et al., 2006). The r.m.s.d. for all heavy backbone atoms in pair-wise comparisons of the different models ranged from 0.7 to 1.2 Å and those for all atoms ranged from 1.8 to 2.2 Å, showing a well-defined global fold. The quality of the models was checked using PROCHECK (Laskowski et al., 1993). For the best-scoring model, 88% of the residues were in the most favored regions of the Ramachandran plot, and the remaining 12% were in additional allowed regions (Laskowski et al., 1993).

2.4. Small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS) experiments

SANS data were collected at the Laue Langevin Institute (Grenoble, France) on beamline D22 (2005). The sample-to-detector distance was 4 m and the neutron wavelength was 0.6 nm so as to cover a scattering vector (Q) range from 0.18 to 1.20 nm^{-1}, where $Q = 4\pi(\sin\theta)/\lambda$. Samples were loaded in a 1 mm pathlength quartz cuvette. The protein concentration was 6.3 mg ml^{-1}. Reduction of SANS data was performed with a program suite available at D22.

SAXS data were collected at the European Synchrotron Radiation Facility (Grenoble, France) on beamlines ID2 (Narayanan et al., 2001) and ID14-3. The sample-to-detector distance was 1 m and the wavelength of the X-rays was 0.0995 nm. Samples were contained in a 1.9 mm wide quartz capillary. The time of exposure was optimized for reducing radiation damage. Data acquisition was performed at 20 °C. Protein concentrations ranged from 1.1 to 6.7 mg ml^{-1} for Par27 and from 2.1 to 8.4 mg ml^{-1} for the PPIase domain. Reduction of the 2D-SAXS patterns to 1D scattering profiles was performed using the established procedure available at ID2, and buffer background runs were subtracted from sample runs.

The radius of gyration and forward intensity at zero angle ($I(0)$) were determined with the programs PRIMUS (Konarev et al., 2003) by using the Guinier approximation at low Q values, in a $Q{\cdot}R_g$ range up to 1.3:

$$\ln I(Q) = \ln I(0) - \frac{R_g^2 Q^2}{3} \quad (1)$$

For SANS experiments, the molecular mass of Par27 was directly calculated from the forward intensity, $I(0)$, knowing the protein concentration and the scattering length density (Jacrot and Zaccai, 1981). For SAXS experiments, the forward scattering intensity was calibrated using bovine serum albumin and lysozyme as references. The radius of gyration and pairwise distance distribution function, $P(r)$, were calculated with the program GNOM (Semenyuk and Svergun, 1991). The maximum dimension (D_{max}) value was adjusted so that the R_g value obtained from GNOM agreed with that obtained from the Guinier analysis.

Models of full-length Par27 and its PPIase domain were obtained by using the program BUNCH (Petoukhov and Svergun, 2005). For the PPIase domain, the all-atom model obtained by template-based modeling was completed by adding dummy residues in place of the 17 N-terminal residues. For full-length Par27, rigid-body modeling was carried out using the X-ray structure of the core domain encompassing the NTD and CTD domains of Par27 and the structure of the PPIase domain predicted by template-based modeling. The NTD and CTD of the core domain from the crystal structure were used as independent sub-domains, thus allowing small fluctuations in their relative orientation. Amino acids 90–99 and 189–195 were replaced by dummy atoms. Modeling of full-length Par27 was performed assuming twofold symmetry around the NTD dimerization interface. Eighty independent models were built, of which 45 were pseudo-enantiomorphs of the crystal structure and were discarded. The normalized spatial discrepancy (NSD) values were calculated with the program DAMAVER between each pair of models without realignment, to evaluate their resemblance and select the most typical model; *i.e.*, the model with the lowest deviation from the others (Table 3). Models were discarded if their NSD value was larger than two standard deviations.

2.5. Molecular dynamics simulation

An all-atom model was constructed from the best BUNCH model of Par27. Using the C_α coordinates generated by BUNCH, the backbone and side chains for the N-terminal His-tag (aa 1–11) and the flexible linkers joining the PPIase domain to the NTD (aa 114–123) and CTD domains (aa 212–219) were built using the SABBACv1.2 server (Maupetit et al., 2006). The resulting all-atom model was energy minimized and used as input for molecular dynamics simulations. Molecular dynamics refinement of the all-atom model of Par27 was carried out using the GROMACS 4.0.2 software package (Hess et al., 2008) and the GROMOS96 53A6 force field (Oostenbrink et al., 2005). The protein was solvated with 27006 SPC/E water molecules in a triclinic box. The minimum distance between any atom of the protein and the box wall was 0.9 nm. Na^+ and Cl^- ions were added to a concentration of 150 mM in order to approximate physiological conditions. Covalent bonds in the protein were constrained using the LINCS algorithm. The geometry of the water molecules was constrained using the SETTLE algorithm. Electrostatic interactions were calculated using the Particle Mesh Ewald (PME) algorithm with a cut-off of 0.9 nm. Lennard–Jones interactions were calculated with a 0.9–1.4 nm twin-range cut-off. Hydrogen atoms were treated as virtual sites to increase computational efficiency. A dielectric constant of 1 and a time step of 5 fs were used. The temperature and the pressure were maintained at 300 K and 1 bar by using the v-rescale thermostat and the Parrinello–Rahman barostat, respectively. The relaxation times were 0.1 and 2.0 ps, respectively. The energy of the system was initially minimized by 2500 steps of steepest descent, followed by 125 ps of position restraints molecular dynamics to let the solvent equilibrate around the protein. Production simulation was run for 20 ns and snapshots were collected every 5 ps for analysis. Root-mean square deviation (r.m.s.d.) and Root-mean square fluctuation (r.m.s.f.) were calculated on the C_α

Table 3
Normalized spatial discrepancy (NSD) and χ values (fit to the experimental SAXS curve) for Par27 models built with BUNCH – 10 best-fitting models. The average NSD value for all pair-wise comparisons is equal to 1.1 and the standard deviation is equal to 0.4.

Models	Consensus models									Outlier models
	8	7	2	6	1	5	9	4	3	10
⟨NSD⟩ before realignment	0.951	0.966	0.976	0.977	0.979	0.988	1.034	1.036	1.001	2.374
NSD after realignment	Ref.	0.760	0.762	0.751	0.860	0.678	0.909	0.570	0.868	–
χ Value	0.51	0.51	0.50	0.51	0.48	0.60	0.53	0.54	0.45	0.56

atoms after least-square fitting of the backbone using the GROMACS routines.

2.6. PPIase activity assays

Two PPIase activity assays were used to characterize the enzymatic activity of the isolated PPIase domain, using either reduced carboxymethylated RNAse T1 or a 16-mer peptide as substrates as previously described for the full-length protein (Hodak et al., 2008). For the peptide assay, NMR experiments were performed on a Bruker 600 MHz DMX spectrometer (Bruker, Karlsruhe, Germany) equipped with a cryogenic triple resonance probe head. All ^{1}H spectra were calibrated with 1 mM sodium 3-trimethylsilyl-d(3,3′,2,2′)-propionate as a reference. Standard TOCSY and NOESY spectra with 60 and 400 ms mixing times, respectively, were acquired on a 2 mM peptide solution in 25 mM phosphate buffer (pH 6.6), 50 mM NaCl at 293 K to assign peptide ^{1}H resonances. The content of *cis* conformers was calculated with the integration of the Hα/Hγ correlation peaks for both *cis* and *trans* signals in a homonuclear TOCSY experiment. To measure the PPIase activity, a ^{1}H–^{1}H EXSY spectrum was recorded at a mixing time of 300 ms on a 2 mM peptide sample containing 0.050 mM (2.5%) Par27 in 25 mM phosphate buffer (pH 6.6), 50 mM NaCl in 100% D_2O. A series of ^{1}H–^{1}H EXSY spectra were then acquired at various mixing times (100, 200, 300, 400 and 500 ms) to quantify the exchange rate of the Par27 isomerase activity. Integration of the Hα *cis*/ Hα *trans* cross-peaks detected for both proline residues was performed. The exchange cross-peaks were normalized with the integration value of the corresponding diagonal peak to provide the ratio of molecules that underwent exchanges from the *cis* to the *trans* conformation during the mixing time. The exchange rate constant (k_{exch}) was calculated by fitting the theoretical curve obtained with the following equation to the experimental points:

$$\%(cis \rightarrow trans) = \alpha \frac{1 - \exp(-(1 + 1/\alpha)k_{exch}M_T}{1 + \exp(-(1 + 1/\alpha)k_{exch}M_T} \quad (2)$$

where $\%(cis \rightarrow trans)$ corresponds to the fraction of molecules that underwent changes from the *cis* to the *trans* conformation during the mixing time (M_T), and α is the excess of *trans* over *cis* conformer.

RNase T1 was purchased from Sigma. The reduction and carboxymethylation of RNase T1 to generate RCM-T1 were performed as previously described (Mücke and Schmid, 1994). RCM-T1 was unfolded in 0.1 M Tris–HCl (pH 8) for at least 1 h at room temperature at a final concentration of 20 μM. The refolding was initiated by a 40-fold dilution in 2 M NaCl. The time course of refolding of the RCM-T1 was followed at 320 nm with a QM-4 spectrofluorometer (PTI) using λ_{exc} = 268 nm. For the calculation of k_{cat}/k_m, Par27 or the PPIase domain were used at a fixed concentration and RCM-T1 concentration was varied from 100 nM to 1 μM. The reaction was measured in a total volume of 2 ml, in a quartz cuvette under constant agitation at 15 °C. For each RCM-T1 concentration, the initial velocity value measured in the absence of enzyme was subtracted from the initial velocity value measured in its presence.

3. Results

3.1. X-ray diffraction analysis of full-length Par27

Our initial goal was to solve the structure of Par27 using X-ray diffraction techniques. Full-length Par27 was crystallized as reported previously (Wohlkönig et al., 2008). Data were collected for a SeMet derivative, indexed in space group C222 and subjected to either the MAD or MIR method for phasing. Surprisingly, only a partial model corresponding to residues 1–101 and 208–239 could be built using automated procedures. A high quality electron-density map was obtained at 2.2 Å resolution for residues 1–89 and 218–239 and the sequence assignment in this part of the structure was straightforward (Fig. 1a). In addition, the backbone structure could be built for residues 90–101 and 208–217, but no electron density could be observed for residues 102–207 that encompass the previously identified PPIase domain (aa 112–199) (Hodak et al., 2008).

To confirm that residues 102–207 were present in the crystal, we dissolved the crystal and verified protein integrity by mass spectrometry, which yielded the expected mass for the full-length protein. The part of the protein missing in the structure was not expected to be intrinsically unstructured, since it includes a clearly identified PPIase domain of the parvulin family, and it was not expected to be flexible, since it appears to be involved in the crystal packing. However, examination of the diffraction patterns revealed the presence of a pronounced diffuse scattering and of many diffraction spots characteristic of an incommensurable lattice. Inspection of experimentally phased electron-density maps confirmed the absence of electron density for the PPIase domain. These results suggested that, within the crystal, the PPIase domain adopts periodically shifted conformations differing from the crystal periodicity, thus smearing out the electron density corresponding to this domain. The void left in the lattice by the absence of this part of the protein was large enough to accommodate the missing PPIase domains in a conformation similar to that obtained in the model reconstructed from SAXS data (see below). No further attempt was made to carry on further this analysis because of the inherent difficulties encountered in collecting, integrating and interpreting diffuse scattering data (Wall et al., 1997; Wilson and Brunger, 2003) and because of our inability to obtain crystals of sufficient quality for this kind of study.

3.2. The NTD/CTD core structure of Par27

The partial crystal structure and sequence alignments (Hodak et al., 2008) indicated that Par27 can be divided into an N-terminal sub-domain (NTD), a central PPIase domain and a C-terminal sub-domain (CTD) (Fig. 1b). In the crystal structure, Par27 forms symmetrical homodimers in which the NTD and CTD sub-domains of both monomers interact to form a cradle-shaped NTD/CTD core structure (Fig. 1c). The NTD of each monomer consists of five α-helices (α1–α5) and a small anti-parallel β-sheet formed by two short β-strands (β1 and β2). The CTD forms a long α-helix

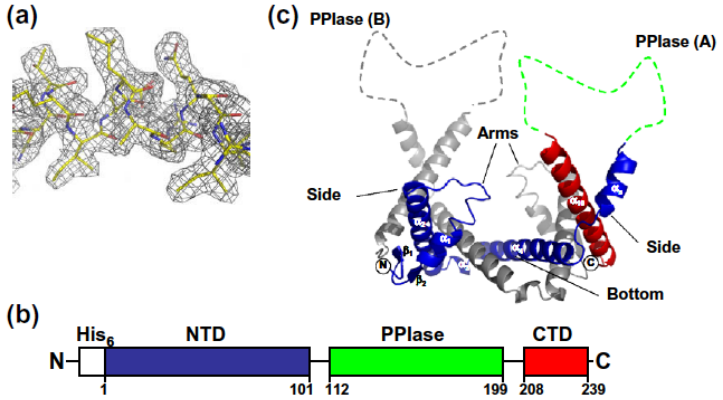


Fig. 1. Crystal structure of the Par27 NTD/CTD core domain. (a) Excerpt of the experimental electron density map of Par27 contoured at 1.4 σ level. The illustrated region is part of helix α3. (b) Schematic diagram of the modular organization of Par27. The His-tag sequence is shown in white, the N-terminal sub-domain (NTD) in blue, the PPIase domain in green and the C-terminal sub-domain (CTD) in red. Amino acid numbers showing the boundaries of the different domains refer to the amino acid sequence of the mature protein, after cleavage of the signal peptide. (c) Ribbon drawing of the Par27 NTD/CTD core dimer showing one protomer with the same color-code as in (a) and the other in gray. The secondary structure elements are labeled in the colored protomer. The dotted lines show the expected locations of the PPIase domains that could not be solved. (For interpretation of color mentioned in this figure the reader is referred to the web version of the article.)

(α10; secondary structure elements have been numbered from the N-terminus to the C-terminus in the final model of the entire protein) that packs against helices α4 and α5 of the NTD in the same monomer and against helix α2 of the NTD in the other monomer (Fig. 1c). In each monomer, the anti-parallel β-strands and helices α1 and α2 form the 'arms', helices α5 and α10 (CTD) form the 'sides' and helices α3 and α4 form a flat dimeric 'bottom' (Fig. 1c).

A sequence homology search using BlastP identified homologues of Par27 mainly in the *Burkholderiales* order (Hodak et al., 2008). However, a search of the Protein Data Bank with the dimeric core domain of Par27 using DALI (Holm and Sander, 1996) revealed structural similarity with the core domain of the periplasmic PPIase/chaperone SurA from *E. coli* (z-score = 6.1) (Bitto and McKay, 2002; Xu et al., 2007). The core domain of SurA also shares structural similarity (Stirling et al., 2006) with the C-terminal clamp-like binding domain of the cytosolic PPIase/chaperone trigger factor (TF) (Ferbitz et al., 2004; Maier et al., 2005) and with the hypothetical *Mycoplasma* protein MPN555, identified as a putative homolog of TF from *Pseudomonas putida* although it contains no PPIase domain (Schulze-Gahmen et al., 2005). These three proteins share with Par27 a modular organization in structural or functional domains although the number of domains and their organization along the amino acid sequence are different (Fig. 2a). The monomeric SurA comprises two PPIase domains of the parvulin family, PPIase 1 and 2, flanked by the N-terminal (NTD) and C-terminal sub-domains (CTD). The NTD, PPIase 1 and CTD domains form a platform to which the enzymatically active PPIase 2 domain is tethered by long extended chains (Bitto and McKay, 2002). TF comprises an N-terminal ribosome binding domain, a central PPIase domain of the FKBP family and two C-terminal ARM domains that binds to nascent proteins leaving the ribosome exit tunnel (Ferbitz et al., 2004; Merz et al., 2006). The MPN555 protein comprises only NTD and CTD sub-domains, but no PPIase domain.

Sequence alignment of the N- and C-terminal sub-domains of Par27 revealed overall 19% identity and 40% similarity with the corresponding sub-domains of SurA (Fig. S1 in Supplementary material), but only 10% similarity with the corresponding domains of TF and MPN555 (data not shown). Although Par27 forms dimers, the topological organization of its core domain is reminiscent of those of the core domains of SurA, TF and MNP555 (Fig. 2b). Comparison of the topological organization of Par27 with that of SurA (Fig. 2b) indicated conservation of strands β1 and β2 and of helices α1 and α2. In SurA, helices α3 and α4 form a close to 90° angle that allows helices α5 and α6 to dock against helices α1, α2 and α18, while in dimeric Par27, the corresponding residues form a long helix, α4, that interacts with helix α10 of the same chain and with helix α2′ of the other chain. This geometry of the polypeptide chain introduces a sub-domain swapping in Par27 involved in the dimerization of the protein mainly through interactions between helices α4 and α4′ that lie anti-parallel to each other in the dimer, but also through contacts between residues of helices α2 and α10′, α2′ and α10, α3 and α4′ and α3′ and α4. The long α-helix of the C-terminal domain in both Par27 (α10) and SurA (α18) and the following short β-strand of SurA (β11) dock onto strand β1 and helix α5. Alignment of the three-dimensional structures using the SSM server (Krissinel and Henrick, 2004) indicated that the anti-parallel β-sheet and helices α1, α2, α3, α5 and α18 of SurA can be partially superimposed on the corresponding secondary structure elements of Par27 (Fig. 2c). An r.m.s.d. of 3.6 Å was obtained for the backbone atoms, suggesting a good similarity between the two proteins, although SurA is monomeric and Par27 dimeric. A SurA variant lacking the second PPIase domain forms a dimer in the presence of a dodecapeptide substrate (Xu et al., 2007), but its topological organization is different from that of Par27 (Fig. 2b and Fig. S2 in Supplementary material). In the dimer of the SurA variant, the main interchain interactions are between helices α18 and α18′ that lie parallel to each other and between helix α6 and helices α3′ and α5′ (as well as α6′ and α3 and α5). In TF, the topological arrangement of helices α7–α11 forming the two arms is similar to that found in SurA (Fig. 2b). However, the long helix α6 is located upstream of the ARM domains in the TF amino acid sequence, although it occupies a position in space similar to that of the long CTD helix of SurA. In the MPN555 protein, the topology is similar to that of SurA except for the presence of an additional β-sheet, formed by stands β1 and β3, and an additional helix, α3, inserted in the sequence at a position equivalent to the PPIase domains of SurA and Par27 (Fig. 2b and Fig. S2 in Supplementary material). As for SurA, the three-dimensional structure of the Par27 NTD/CTD core domain can be aligned with corresponding regions of TF and MPN555 (Fig. S2 in Supplementary material).

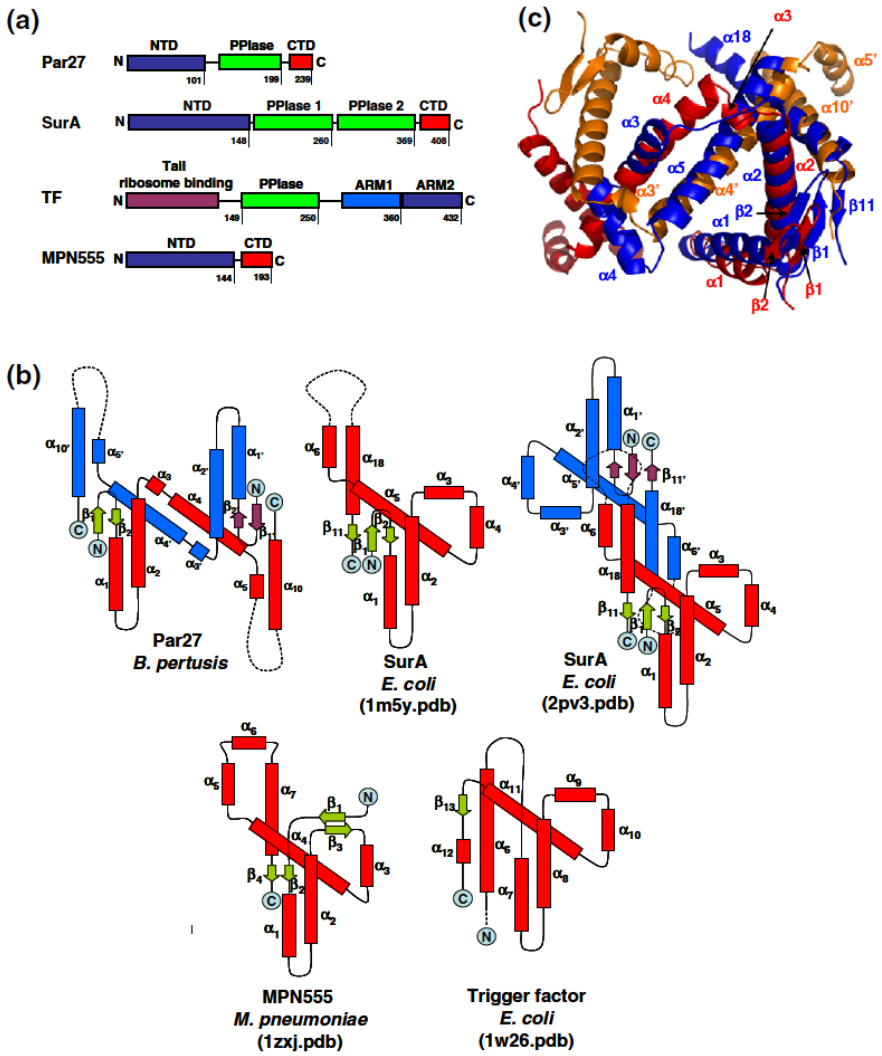


Fig. 2. Structural comparison of the Par27 core with other clamp-like PPIase/chaperones. (a) Domain organization of *B. pertussis* Par27 compared with that of other clamp-like PPIase/chaperones: SurA and trigger factor from *E. coli* and MPN555 protein from *M. pneumoniae*. The numbers indicate the C-terminal boundaries of the different domains in the mature proteins. (b) Topology of the core domains of the different proteins. The green arrows and red boxes show β-strands and α-helices, respectively. In Par27 and SurA dimers, violet arrows and blue boxes show β-strands and α-helices in the second protomer. The secondary structure elements in the second protomer are labelled with a prime. (c) Superposition of the Par27 NTD/CTD core dimer with the SurA core domain (NTD and CTD sub-domains). Superposition of the structure was performed with SSM (Krissinel and Henrick, 2004). The α carbon r.m.s.d. was 3.6 Å. SurA is shown in blue, one Par27 protomer in red and the other in orange. The corresponding secondary structure elements are labeled. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

3.3. Small-angle scattering experiments

In order to build a structural model of the entire protein, small-angle X-ray (SAXS) and neutron (SANS) scattering data were combined with molecular modeling and with the crystal structure of the NTD/CTD core (Putnam et al., 2007; Tsutakawa et al., 2007). The PPIase domain and full-length Par27 were produced as recombinant proteins with an N-terminal His-tag and purified as previously described (Hodak et al., 2008). Circular dichroism spectroscopy revealed that both proteins were folded. The monodispersity of all samples used in SAXS and SANS experiments was checked by size exclusion chromatography (SEC) combined with detection by multi-angle laser light scattering (MALLS) and refractometry (RI) (Gérard et al., 2007; Hodak et al., 2008). The analysis of the isolated PPIase domain by SEC-MALLS-RI yielded a molecular mass of 12.0 ± 0.5 kDa, very close to the calculated mass, indicating that it is monomeric (Table 2).

SAXS profiles for full-length Par27 and the isolated PPIase domain were collected for scattering vector values ($Q = 4\pi(\sin\theta)/\lambda$) ranging from 0.1 to 3.0 nm^{-1} and from 0.2 to 2.3 nm^{-1}, respectively (Figs. 3a and 4a). A SANS profile was also collected for full-length Par27 (Fig. 4a). Guinier plots were linear for $Q{\cdot}R_g$ values ranging from 0.4 to 1.3 and yielded radius of gyration (R_g) values independent of protein concentrations (Figs. 3b and 4b and Table 2) (Glatter and Kratky, 1982). Similar R_g values were obtained from the distance distribution function, $P(r)$, calculated as the inverse Fourier transform of $I(Q)$ using the program GNOM (Semenyuk and Svergun, 1991) (Figs. 3c and 4c and Table 2). The

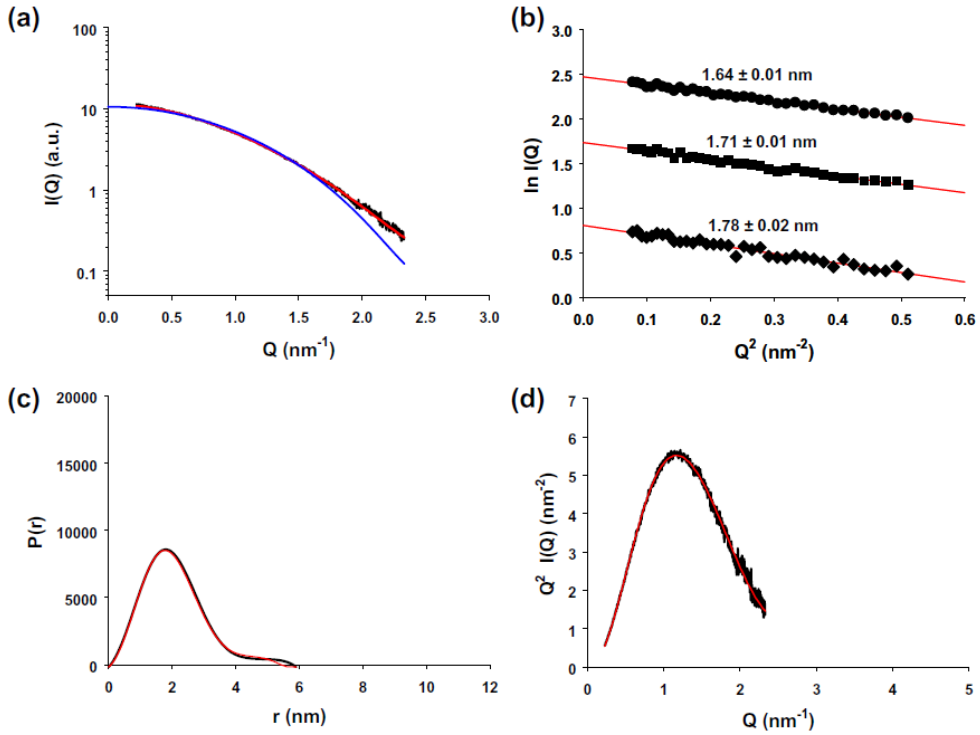


Fig. 3. SAXS profiles of the Par27 PPIase domain. (a) Experimental SAXS data as a function of Q. The black line shows the experimental profile, the blue line the SAXS profile calculated for the best-scoring template-based model (χ value = 4.40) and the red line the curve calculated for the reference model including the N-terminal extension generated with BUNCH (χ value = 0.78). (b) Guinier plots for Q values ranging from 0.22 to 0.73 nm^{-1} ($0.4 < Q \cdot R_g < 1.2$). The symbols are for data measured at protein concentration of 8.4 mg ml^{-1} (circles), 4.2 mg ml^{-1} (squares) and 2.1 mg ml^{-1} (diamonds). The red lines show linear regressions, and the numbers above the curves show the R_g values calculated from the slope of the fitted curves. (c) Distance distribution function and (d) Kratky plot. The $P(r)$ function was calculated using the GNOM software and a D_{max} value of 5.9 nm. The experimental curve is shown in black and the theoretical curve calculated for the reference model generated with BUNCH in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

forward scattering intensity, $I(0)$, provided estimates of the respective molecular masses, in agreement with the calculated molecular masses for the dimeric Par27 and the monomeric PPIase domain and with the molecular masses measured by multi-angle laser light scattering (MALLS) ((Hodak et al., 2008) and this work). These results confirmed that Par27 is dimeric, that the PPIase domain is monomeric and that both proteins are well-behaved and monodisperse in solution.

The $P(r)$ function shows the probability distribution of interatomic distances within the scattering particle. For the PPIase domain, the pair distribution function, $P(r)$, (Fig. 3c) has a symmetrical bell shape typical of a globular protein with a peak centered near 2 nm indicating a maximal dimension of about 4 nm, and the Kratky plot shows a characteristic symmetrical peak, typical of a folded protein (Doniach, 2001; Svergun and Koch, 2003) (Fig. 3d). In the $P(r)$ function, the bell shape curve is followed by a tail extending to a maximum protein dimension (D_{max}) of 5.9 nm. A spherical protein of this size (12.2 kDa, $\bar{v} = 0.73$ cm^3 gr^{-1}) is expected to have a diameter of about 3.0 nm, suggesting that the PPIase domain is slightly elongated, whereas the N-terminal His-tag could explain the tailing. For the full-length Par27, both the $P(r)$ function (Fig. 4c) and the Kratky plot (Fig. 4d) show complex profiles with several maxima and shoulders. In the $P(r)$ function, the first peak near 2 nm might correspond to the distribution of intra-domain distances, while the other peak centered roughly around $D_{max}/2$ (5.5 nm), might correspond to the distribution of inter-domain distances. The linear rise at high Q values observed in the Kratky plot ($Q > 3$ nm^{-1}) suggests the presence of flexible regions (Doniach et al., 1995; Doniach, 2001).

3.4. Modeling the structure of the PPIase domain of Par27

The PPIase domain of Par27 (aa 112–200) belongs to the parvulin family (Hodak et al., 2008). Pairwise sequence alignments revealed between 35% and 40% identity with six PPIase domains of known three-dimensional structure and multiple sequence alignment revealed several conserved elements (Fig. S3a in Supplementary material) (Ranganathan et al., 1997; Sekerina et al., 2000; Terada et al., 2001; Bitto and McKay, 2002; Bayer et al., 2003; Kühlewein et al., 2004; Li et al., 2005; Tossavainen et al., 2006). A high-resolution structural model was obtained by template-based modeling by using the LOMETS metaserver (Wu and Zhang, 2007) (Fig. S3b in Supplementary material).

The scattering curve calculated from the best-scoring model using the program CRYSOL (Svergun et al., 1995) show, however, discrepancies at low and high Q values, and the quality of the fit to the experimental curve is poor as judged by a χ value of 4.4 (Fig. 3a). The purified protein used in the SAXS experiments comprises a 17-residues N-terminal extension, including a His-tag and five residues of Par27 that were not included in the homology modeling procedure (MRGSHHHHHHGSGKMEY). Twenty independent models of the entire construct were thus generated with the program BUNCH (Petoukhov and Svergun, 2005) by adding dummy residues at the N-terminus of the high-resolution structure obtained

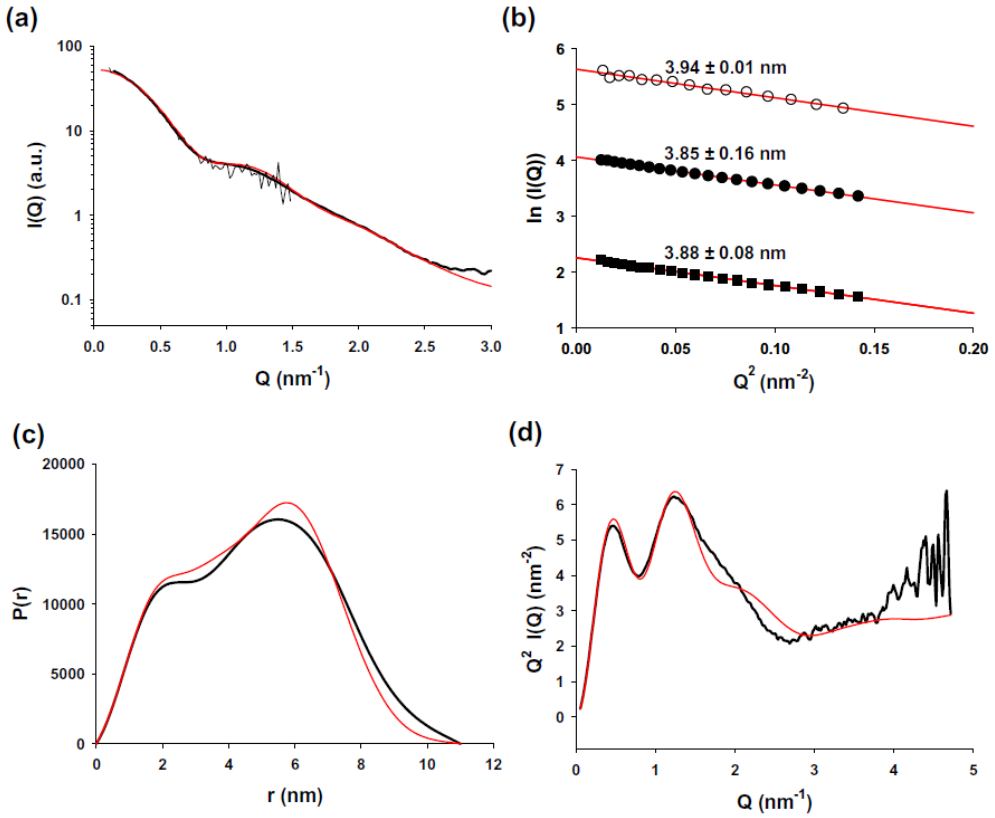


Fig. 4. SAXS and SANS profiles of full-length Par27. (a) Experimental SAXS and SANS data as a function of Q. The thick black line shows the experimental SAXS profile, the thin black line the SANS profile and the red line the theoretical curve calculated for the averaged BUNCH models (χ value = 0.50). (b) Guinier plots for Q values ranging from 0.11 to 0.37 nm^{-1} ($0.4 < Q.R_g < 1.4$). The open circles are for SANS data measured at a protein concentration of 6.1 mg ml^{-1}. The closed symbols are for SAXS data measured at protein concentration of 6.7 mg ml^{-1} (circles) and 1.1 mg ml^{-1} (squares). The red lines show linear regressions, and the numbers above the curves are the R_g values calculated from the slope of the fitted curves. For clarity, each plot has been shifted along the ln($I(Q)$) axis. (c) Distance distribution function and (d) Kratky plot. The $P(r)$ function was calculated using the GNOM software and a D_{max} value of 11.5 nm. The experimental curve is shown in black and the theoretical curve calculated for the averaged nine BUNCH models in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

by homology modeling. After alignment of the structured domain using the program DAMAVER (Volkov and Svergun, 2003) (average normalized spatial discrepancy value, ⟨NSD⟩ = 0.48), the added tail exhibits a different conformation in the various models suggesting that it is flexible (Fig. 5). The globular domain is a prolate spheroid with a maximum dimension of 3.9 nm, in good agreement with the SAXS curve analysis, and the presence of the His-tag explains the large dimension of the molecule (D_{max} = 5.9 nm) (Fig. 3). The inclusion of dummy residues corresponding to the N-terminal extension significantly improved the quality of the fit to the experimental curve, yielding χ values for the 20 different models that range from 1.41 to 1.43 (Fig. 3a).

3.5. Rigid-body modeling of full-length Par27 structure

Models of the full-length protein were generated by rigid-body modeling with the program BUNCH (Petoukhov and Svergun, 2005) using the crystal structure of the dimeric NTD/CTD core and the template-based model of the PPIase domain. Because building a three-dimensional structure from small-angle scattering data is a problem with more than one solution, 35 independent models were compared. Fig. 6a shows a typical model of Par27 where the PPIase domains are tethered to the dimeric NTD/CTD core through two shafts, each one formed by helix α5 from the NTD of one monomer and helix α10 from the CTD of the other monomer. The theoretical curves computed for the 35 models yielded χ values that ranged from 0.45 to 1.14. Pair-wise comparison using the program DAMAVER (Volkov and Svergun, 2003) yielded an average normalized spatial discrepancy value (⟨NSD⟩) of 1.12 (standard deviation = 0.31). Two models were discarded because their NSD values were larger than two standard deviations. If only the best-fitting models with χ values lower or equal to 0.6 were selected (10 models), pair-wise comparison using the program DAMAVER yielded a similar ⟨NSD⟩ value of 1.14, and the individual NSD values calculated after realignment of each model to the reference model ranged from 0.57 to 0.91 (Table 3). For globular proteins, NSD values larger than 0.7 are indicative of some variability among the models (Volkov and Svergun, 2003). Superposition of the models revealed that they differ mainly in the orientation of the PPIases domains by rotation around the shafts (Fig. 6b). To account for that flexibility, the 10 best-fitting curves were averaged. The averaged curve correctly reproduces the experimental curve (χ value = 0.50) (Fig. 4a) as well as the calculated distance distribution function (Fig. 4c) and Kratky plot (Fig. 4d). Although this procedure provided no improvement to the quality of the fit, an ensemble of conformations appears more relevant for representing a flexible protein.

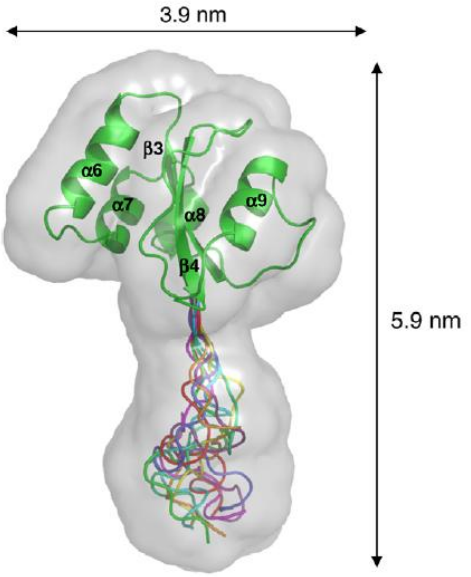


Fig. 5. Solution structure of the Par27 PPIase domain. BUNCH models of the PPIase domain with an N-terminal 6-His-tag. The template-based model of the PPIase domain is shown in ribbon representation and the secondary structure elements are labeled. The extended 17-residues N-terminal extensions from 10 BUNCH models are shown as ribbons of different colors. The gray surface shows the filtered model generated by the program DAMAVER (Damfilt model). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

To test how SAXS data constrain the positioning of the PPIase domains, models were built by manually moving the PPIase domains relative to the dimeric NTD/CTD core, and the calculated SAXS profiles were compared with the experimental data. Models in which the PPIase domains were moved apart (Fig. 7b and f) or toward (Fig. 7d) the NTD/CTD core clearly show poorer fit of the calculated and experimental SAXS profiles (Fig. 7a, c and e), stating that SAXS data are adequate for positioning the PPIase domains within the protein.

3.6. Dynamics in the structure of full-length Par27

To analyze the conformational fluctuations occurring in Par27, a 20 ns-long molecular dynamics simulation was run in explicit solvent and the conformational drift of the structure was measured in terms of root-mean-square deviations (r.m.s.d.) (Fig. 8a). After an initial rise during the first 10 ns in which the structure adjusted to the simulation conditions, the r.m.s.d. value reached a plateau, characteristic of a stable structure. The location of the flexible parts of the protein was determined by calculating the root-mean-square fluctuations (r.m.s.f.) of C_α atoms with respect to their time-averaged position over the last 10 ns of the simulation (Fig. 8b). The plot of the r.m.s.f. along the amino acid sequence revealed largest movements of helix α1, the C-terminal part of helix α5 and the PPIase domains (Fig. 8c and d) in agreement with the BUNCH models.

3.7. Functional characterization of the Par27 domains

The PPIase activity of full-length Par27 was previously characterized using two different assays, either by monitoring the catalyzed refolding of reduced and carboxymethylated RNAse T1

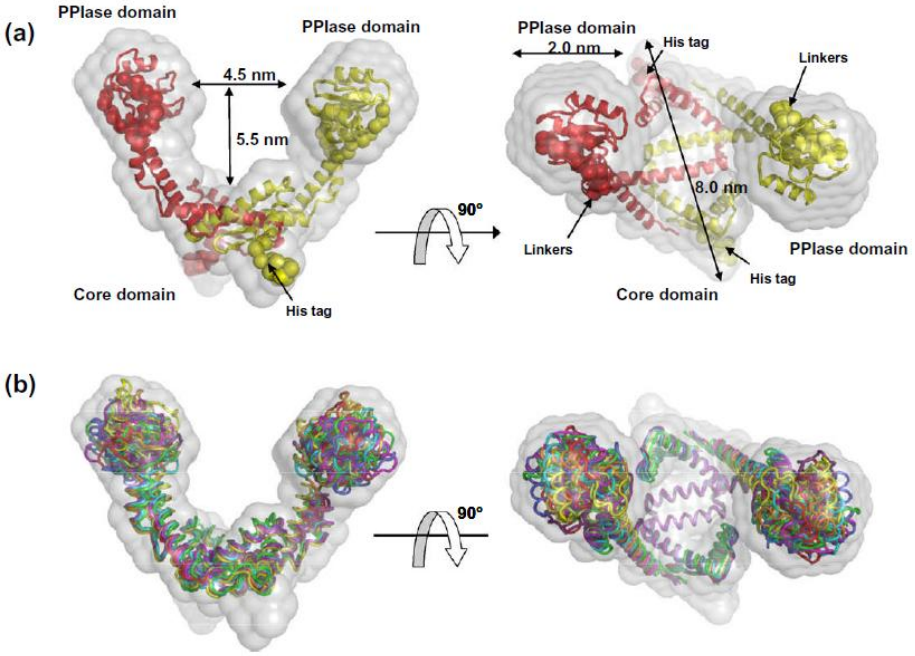


Fig. 6. Solution structure of Par27 and its domain organization. (a and b) Model generated by rigid-body modeling of the SAXS data using BUNCH. The ribbon representations depict the 'core' domain determined by X-ray crystallography and the PPIase domains predicted by homology modeling. The beads show the dummy residues added by BUNCH corresponding to the N-terminal His-tag and the two linkers tethering the PPIase domains to the NTD/CTD core.

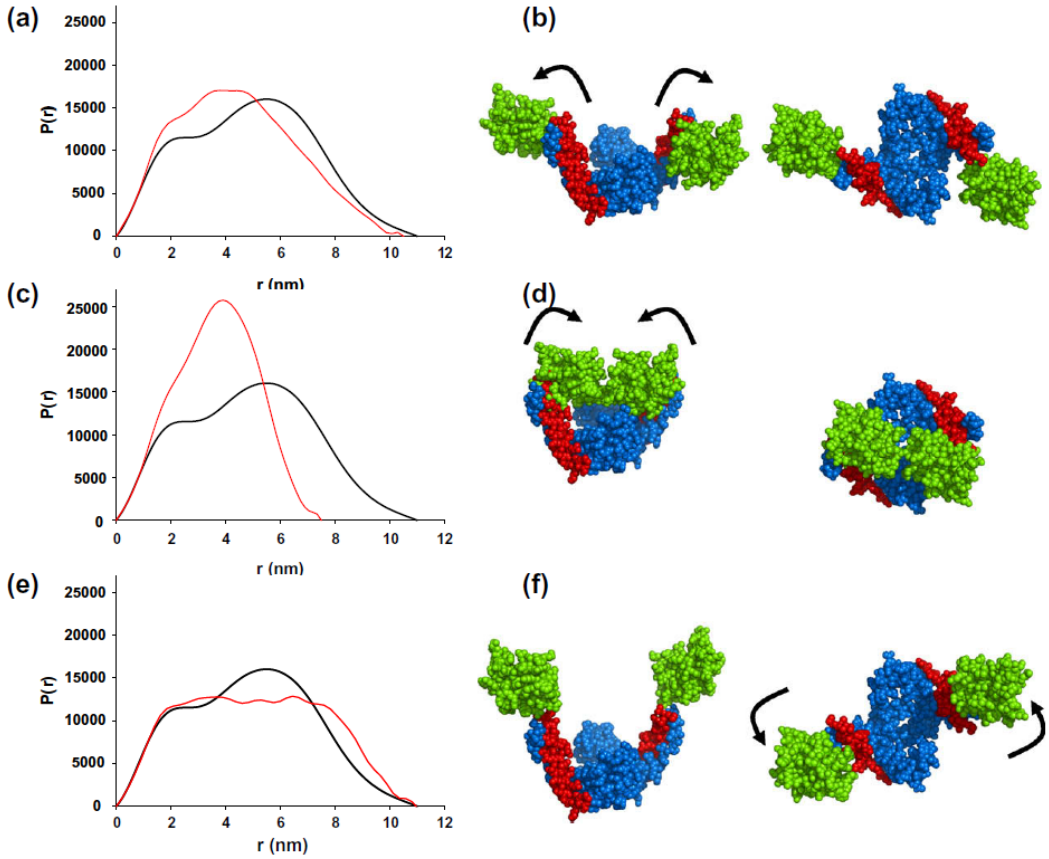


Fig. 7. Models of Par27 and $P(r)$ functions. (b, d, and f) The models were generated by moving the PPIase domains relative to the NTD/CTD core. (a, c, and e) The $P(r)$ function was calculated for the different models and compared with the experimental curve. The experimental curve is shown in black and the calculated curve in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

(RCM-T1), or by measuring the isomerisation rate of two peptidyl-prolyl bonds in a 16-mer peptide (Hodak et al., 2008). The activity of the isolated PPIase domain was significantly lower than that of full-length Par27 in both tests, and in particular the activity with the protein substrate was strongly affected. The k_{cat}/k_m determined for refolding of RCM-T1 catalyzed by the PPIase domain was reduced more than 140-fold as compared with full-length Par27 (Fig. 9a), whereas the exchange rate measured for the 16-mer peptide was reduced only fivefold (Fig. 9b). These differential behaviors suggested that the NTD/CTD core plays a role in binding large protein substrates or in stimulating the catalytic activity of the PPIase domains. It is noteworthy that no stable complex could be formed between denatured Fha30 and the isolated PPIase domain (data not shown), in contrast with what was observed with full-length Par27 (Hodak et al., 2008), supporting the involvement of the NTD/CTD core in binding protein substrates.

Previously, we showed that Par27 prevents the aggregation of lysozyme and forms a tight complex with Fha30, a 30 kDa fragment of *B. pertussis* haemagglutinin, only when Par27 was mixed with the unfolded form of Fha30 (Hodak et al., 2008). Here, we found that the isolated PPIase domain has no effect on lysozyme aggregation and that no complex between the PPIase domain and unfolded Fha30 could be detected by SEC, indicating that the NTD/CTD platform plays a major role in binding to the client protein.

4. Discussion

Par27 of *B. pertussis* belongs to a large group of bacterial folding catalysts that exhibit both chaperone and PPIase activities. All three families of PPIases – cyclophilins, FKBPs and parvulins – comprise single domain as well as multi-domain proteins. In the parvulin family, Par10 and Par14 are single domain proteins, whereas Pin1, SurA or Par27 contain additional domains. Most often the additional domains act as chaperone and/or mediate the oligomerization of the protein (Saul et al., 2004). Par27 is the first example of a multi-domain protein containing a parvulin PPIase domain that is dimeric (Hodak et al., 2008).

In Par27, the NTD and CTD sub-domains mediate the dimerization, whereas the central domain carries the PPIase activity. CD spectroscopy and SAXS indicate that the isolated PPIase domain is properly folded, apart from the His-tag tail, but this isolated domain exhibits a reduced peptidyl-prolyl isomerase activity both toward a peptide substrate and a protein substrate when compared to the entire Par27. In both SurA and TF, the isolated active PPIase domain exhibits a reduced PPIase activity toward protein substrates but conserves the same activity than the full-length protein when assayed with a peptide substrate (Scholz et al., 1997b; Behrens et al., 2001). This suggests that the dimeric NTD/CTD core domain of Par27 plays a role in maintaining the proper arrangement of the PPIase active site residues and/or participates in substrate

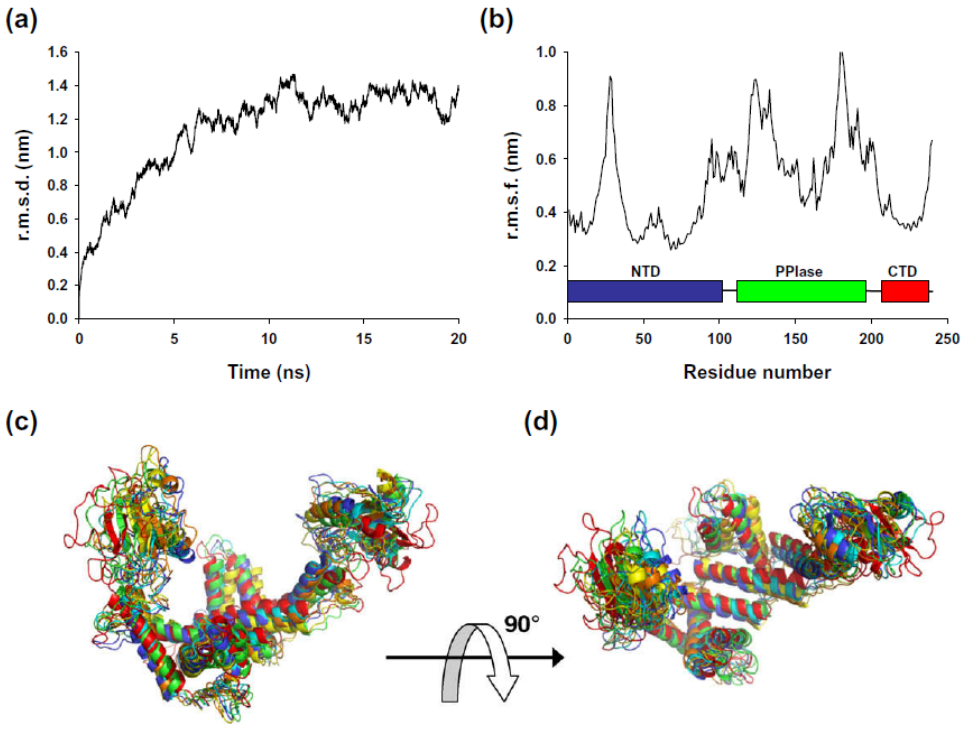


Fig. 8. Molecular dynamics simulation. (a) Root-mean square deviations (r.m.s.d.) of all C_α atoms from the starting structure. (b) Root-mean square fluctuations (r.m.s.f.) of all C_α atoms from their time-averaged positions. The inset shows the boundaries of the different domains and sub-domains of Par27. (c and d) Superposition of snapshots taken at 12, 14, 16, 18 and 20 ns of the MD simulation.

binding and/or optimizes the presentation of the substrate to the catalytic site.

In other multi-domain protein folding catalysts that exhibit both a protein folding catalytic activity and a chaperone activity, the two functions were assigned to distinct regions of the molecule. Structural comparison with these proteins suggests that the cradle-like NTD/CTD core of Par27 resembles the clamp-like structures found in other chaperone proteins (Stirling et al., 2006) and is a good candidate for the substrate-binding site. On one hand, the overall V-shape of the Par27 homodimer (Fig. 6) resembles that of other dimeric chaperones/folding catalysts. FkpA from the periplasm of *E. coli* (Saul et al., 2004), and Mip, a cell surface virulence factor of *Legionella pneumophila* (Riboldi-Tunnicliffe et al., 2001) are chaperone/PPIase protein folding helpers that form V-shaped dimers with the C-terminal PPIase domain tethered to a central dimeric core by a long α-helix. In FkpA deletion studies showed that the chaperone activity resides in the N-terminal dimeric domain (Saul et al., 2004). Another folding catalyst, the disulfide isomerase DsbC from the periplasm of *E. coli* also forms a V-shaped homodimer consisting of separate thioredoxin domains linked to a dimeric core domain by a long α-helix (McCarthy et al., 2000). In DsbC, the core domain formed by the dimerization of NTDs carries the chaperone activity. There is no sequence or structural similarity between the NTD/CTD core of Par27 and the corresponding regions of other dimeric protein folding helpers, apart from the overall V-shape structure, suggesting that these clamp-like structures result from a convergent evolution because they provide a efficient mechanism for seizing large protein substrates, preventing their aggregation, and giving them access to the catalytic domains (Stirling et al., 2006).

On another hand, the dimeric NTD/CTD core domain of Par27 has a topological organization similar to the chaperone substrate-binding domain of monomeric proteins such as SurA and trigger factor. In SurA, the chaperone activity was found to be independent of the presence of the active PPIase 2 domain (Behrens et al., 2001; Bitto and McKay, 2002). The putative substrate-binding site, suggested from intermolecular packing interactions observed in the crystal, is located in the core domain and is made of three helices (α1, α2 and α5) of the NTD, forming a crevice where a short peptide can bind {Bitto, 2002 #4631; Bitto, 2003 #4630}. In TF, the chaperone activity is also independent of the presence of the catalytic PPIase activity (Scholz et al., 1998; Kramer et al., 2004a), and the N-terminal and the two 'arm' domains form a cradle-like structure that was suggested to be important for substrate binding (Ferbitz et al., 2004; Stirling et al., 2006), although the location of the chaperone activity could not be clearly assigned to one of the domains (Kramer et al., 2004b). The topological similarity suggests a common ancestry with these monomeric PPIase/chaperone proteins. The dimerization of Par27 could represent an adaptation to its substrates, potentially providing a larger binding surface for its client proteins. In *B. pertussis*, Par27 shows affinity for different amphipathic β-rich proteins or protein domains of outer membrane proteins (Hodak et al., 2008). *In vitro*, it accepts Fha30, a soluble fragment of FHA, only in its unfolded form but not in its native form (Hodak et al., 2008). The dimensions of the cradle-like groove are 8.0 × 4.5 nm, with a depth of 5.5 nm, sufficient to bind a protein domain or to hide from the solvent a large part of a substrate the size of a partially folded Fha30. The motif recognized in the client protein by Par27 is unknown, but is certainly related to the character of the binding surface. The surface

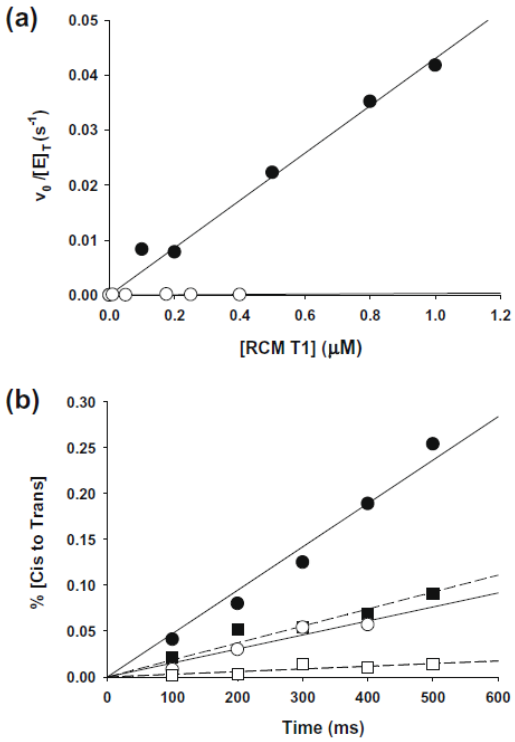


Fig. 9. PPIase activity assays. (a) PPIase activity using an RCM-T1 refolding assay. Refolding of denatured RCM-T1 was initiated by dilution in 2.0 M NaCl. Refolding kinetics was followed by the change of fluorescence at 320 nm at 15 °C as a function of RCM-T1 concentration. Initial velocities were measured in the presence and in the absence of Par27 (filled circles) or of the PPIase domain (open circles). The contribution from the uncatalyzed folding reaction increases linearly with RCM-T1 concentration, and the initial velocities in the presence of Par27 were corrected for the uncatalyzed contribution. Value of 300 ± 100 M^{-1} s^{-1} was obtained for k_{cat}/k_m from the slope of the line. The data for full-length Par27 are taken from (Hodak et al., 2008). (b) PPIase activity using a 16-mer peptide containing two Ser–Pro motifs. The peptide was solubilized at a concentration of 2 mM in 50 mM sodium phosphate (pH 6.6) and 50 mM NaCl. The ratios of molecules that undergo from cis to trans conformation [%(*cis* → *trans*)] for Pro5 (circles) and Pro13 (squares) residues in the presence of 2.5% Par27 are given as a function of time for Par27 (filled symbols) and for the PPIase domain (open symbols). The data for full-length Par27 are taken from Hodak et al. (2008).

of the cradle-like cavity is mainly positive, with a patch of negative potential on each side (Fig. S4a in Supplementary material). Also, this cavity in Par27 contains hydrophobic residues, forming putative hydrophobic pockets in the 'armpits' on each side of the core platform that could accommodate non-polar side chains (Fig. S4b in Supplementary material).

Although Par27 was able to crystallize, the structure of the PPIase domain was not seen in the electron density map, and the structures of the C-terminal part of NTD and the N-terminal part of CTD were not well-defined. These results presented here support a model of Par27 where the PPIase domains are tethered to the NTD/CTD core by flexible linkers and undergo limited motions around the shaft (Fig. 6b). Such flexibility could explain the missing electron density in the crystal structure and could be relevant to the biological activity of the protein. Flexibility appears to be a common feature among the multi-domain chaperone/folding catalysts. In monomeric SurA as well as in dimeric Par27, FkpA, Mip or DsbC, the catalytic domain is tethered to the core by a long α-helix or an extended chain that allows motions between the core and catalytic domains. SurA is stabilized in the crystal by packing interactions with neighboring molecules, however, in the presence of peptide substrates, the protein adapts its structure and binds peptides in different conformations (Xu et al., 2007). Different crystal forms of FkpA and DsbC have also revealed flexibility in the orientation of the C-terminal domains located at the extremities of the V-shape structures (Saul et al., 2004), and the plasticity of FkpA has been confirmed by NMR (Hu et al., 2006). The results reported here for Par27 suggest a similar plasticity that could allow the binding to substrates of different size or the adaptation to substrate size during folding.

Par27 is an abundant periplasmic protein of *B. pertussis* that might serve as a general periplasmic chaperone partially overlapping SurA functions or serving as a back-up to SurA in the periplasm of *B. pertussis* (Hodak et al., 2008). While SurA appears to be essential in *B. pertussis*, the inactivation of the *par27* gene has a small but detectable effect on bacterial growth, indicating that Par27 contributes to bacterial fitness. As for other PPIase/chaperones, an *in vivo* role for the PPIase activity has not yet been demonstrated. In *E. coli*, a SurA variant lacking the active parvulin domain (PPIase 2) complements a null mutation, arguing that the chaperone rather than the PPIase activity of SurA is important *in vivo* (Behrens et al., 2001). It can nevertheless not be ruled out that Par27 also contributes to protein folding in the periplasm by its enzymatic activity.

Acknowledgments

This work was supported in part by a Grant "ANR MIME 07 TPS Path" to F.J.D. The authors thank Dr. H. Drobecq for his help with mass spectrometry analyses, Ambroise Desfosses for his help in writing scripts and Dr. G. Zaccai for his help with SANS experiments. We thank Jean-Paul Declercq, Miguel Ortiz-Lombardia, Christian Cambillau, Javier Perez, Markus Rudolph and Coralie Bompard for helpful discussions. C.L. was supported by a MENRT predoctoral fellowship from the French government. F.J.-D. is a researcher of the CNRS.

Appendix A. Supplementary data

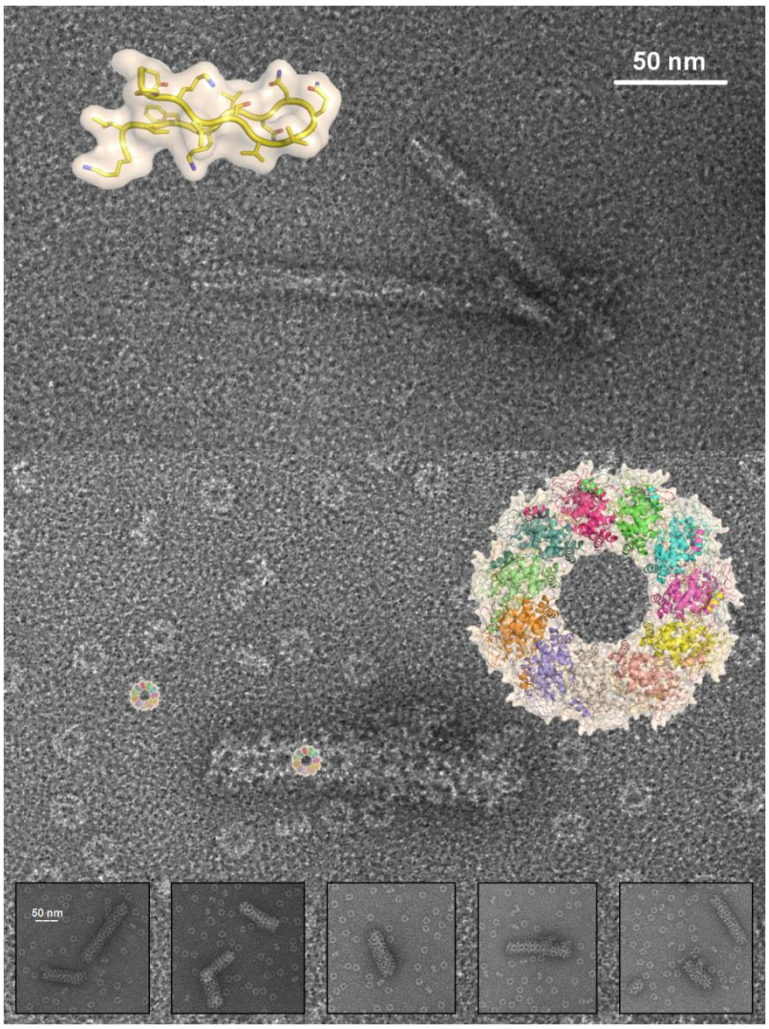
Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jsb.2009.11.007.

References

Bayer, E., Goettsch, S., Mueller, J.W., Griewel, B., Guiberman, E., Mayr, L.M., Bayer, P., 2003. Structural analysis of the mitotic regulator hPin1 in solution: insights into domain architecture and substrate binding. J. Biol. Chem. 278, 26183–26193.
Behrens, S., Maier, R., de Cock, H., Schmid, F.X., Gross, C.A., 2001. The SurA periplasmic PPIase lacking its parvulin domains functions in vivo and has chaperone activity. EMBO J. 20, 285–294.
Bitto, E., McKay, D.B., 2002. Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. Structure 10, 1489–1498.
Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., Paciorek, W., 2003. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallogr. D Biol. Crystallogr. 59, 2023–2030.
Clantin, B., Hodak, H., Willery, E., Locht, C., Jacob-Dubuisson, F., Villeret, V., 2004. The crystal structure of filamentous hemagglutinin secretion domain and its implications for the two-partner secretion pathway. Proc. Natl. Acad. Sci. USA 101, 6194–6199.
Clantin, B., Delattre, A.S., Rucktooa, P., Saint, N., Meli, A.C., Locht, C., Jacob-Dubuisson, F., Villeret, V., 2007. Structure of the membrane protein FhaC: a member of the Omp85-TpsB transporter superfamily. Science 317, 957–961.
Collaborative Computational Project, Number 4, 1994. The CCP4 Suite: Programs for Protein Crystallography, 1994. Acta Crystallogr. D Biol. Crystallogr. 50, 760–763.
Doniach, S., 2001. Changes in biomolecular conformation seen by small angle X-ray scattering. Chem. Rev. 101, 1763–1778.
Doniach, S., Bascle, J., Garel, T., Orland, H., 1995. Partially folded states of proteins: characterization by X-ray scattering. J. Mol. Biol. 254, 960–967.

Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E., Ban, N., 2004. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. Nature 431, 590–596.
Gérard, F., Ribeiro, E., Albertini, A., Zaccai, G., Ebel, C., Ruigrok, R., Jamin, M., 2007. Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution. Biochemistry 46, 10328–10338.
Glatter, O., Kratky, O., 1982. Small Angle X-ray Scattering. Academic Press, London, UK.
Göthel, S.F., Marahiel, M.A., 1999. Peptidyl-prolyl *cis–trans* isomerases, a superfamily of ubiquitous folding catalysts. Cell. Mol. Life Sci. 55, 423–436.
Hess, B., Kutzner, C., Van der Spoel, D., Lindahl, E., 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theory Comput. 4, 435–447.
Hodak, H., Wohlkönig, A., Smet-Nocca, C., Drobecq, H., Wieruszeski, J.M., Senechal, M., Landrieu, I., Locht, C., Jamin, M., Jacob-Dubuisson, F., 2008. The peptidyl-prolyl isomerase and chaperone Par27 of *Bordetella pertussis* as the prototype for a new group of parvulins. J. Mol. Biol. 376, 414–426.
Holm, L., Sander, C., 1996. Mapping the protein universe. Science 273, 595–603.
Hu, K., Galius, V., Pervushin, K., 2006. Structural plasticity of peptidyl-prolyl isomerase sFkpA is a key to its chaperone function as revealed by solution NMR. Biochemistry 45, 11983–11991.
ILL, T.y.b.o., 2005. The yellow book of ILL: a guide to neutron research facilities.
Jacob-Dubuisson, F., Fernandez, R., Coutte, L., 2004. Protein secretion through autotransporter and two-partner pathways. Biochim. Biophys. Acta 1694, 235–257.
Jacrot, B., Zaccai, G., 1981. Determination of molecular weight by neutron scattering. Biopolymers 20, 2413–2426.
Kabsch, W., 1993. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Crystallogr. 26, 795–800.
Kajava, A.V., Cheng, N., Cleaver, R., Kessel, M., Simon, M.N., Willery, E., Jacob-Dubuisson, F., Locht, C., Steven, A.C., 2001. Beta-helix model for the filamentous haemagglutinin adhesin of *Bordetella pertussis* and related bacterial secretory proteins. Mol. Microbiol. 42, 279–292.
Konarev, P.V., Volkov, V.V., Sokolova, A., M.H.J., K., Svergun, D.I., 2003. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J. Appl. Crystallogr. 36, 1277–1282.
Kramer, G., Patzelt, H., Rauch, T., Kurz, T.A., Vorderwulbecke, S., Bukau, B., Deuerling, E., 2004a. Trigger factor peptidyl-prolyl *cis/trans* isomerase activity is not essential for the folding of cytosolic proteins in *Escherichia coli*. J. Biol. Chem. 279, 14165–14170.
Kramer, G., Rutkowska, A., Wegrzyn, R.D., Patzelt, H., Kurz, T.A., Merz, F., Rauch, T., Vorderwulbecke, S., Deuerling, E., Bukau, B., 2004b. Functional dissection of *Escherichia coli* trigger factor: unraveling the function of individual domains. J. Bacteriol. 186, 3777–3784.
Krissinel, E., Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D Biol. Crystallogr. 60, 2256–2268.
Kühlewein, A., Voll, G., Hernandez Alvarez, B., Kessler, H., Fischer, G., Rahfeld, J.U., Gemmecker, G., 2004. Solution structure of *Escherichia coli* Par10: the prototypic member of the Parvulin family of peptidyl-prolyl *cis/trans* isomerases. Protein Sci. 13, 2378–2387.
Laskowski, R.A., Macarthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291.
Lazar, S.W., Kolter, R., 1996. SurA assists the folding of *Escherichia coli* outer membrane proteins. J. Bacteriol. 178, 1770–1773.
Lazar, S.W., Almiron, M., Tormo, A., Kolter, R., 1998. Role of the *Escherichia coli* SurA protein in stationary-phase survival. J. Bacteriol. 180, 5704–5711.
Li, Z., Li, H., Devasahayam, G., Gemmill, T., Chaturvedi, V., Hanes, S.D., Van Roey, P., 2005. The structure of the *Candida albicans* Ess1 prolyl isomerase reveals a well-ordered linker that restricts domain mobility. Biochemistry 44, 6180–6189.
Maier, T., Ferbitz, L., Deuerling, E., Ban, N., 2005. A cradle for new proteins: trigger factor at the ribosome. Curr. Opin. Struct. Biol. 15, 204–212.
Maupetit, J., Gautier, R., Tuffery, P., 2006. SABBAC: online Structural Alphabet-based protein BackBone reconstruction from Alpha-Carbon trace. Nucleic Acids Res. 34, W147–W151.
McCarthy, A.A., Haebel, P.W., Torronen, A., Rybin, V., Baker, E.N., Metcalf, P., 2000. Crystal structure of the protein disulfide bond isomerase, DsbC, from *Escherichia coli*. Nat. Struct. Biol. 7, 196–199.
Merz, F., Hoffmann, A., Rutkowska, A., Zachmann-Brand, B., Bukau, B., Deuerling, E., 2006. The C-terminal domain of *Escherichia coli* trigger factor represents the central module of its chaperone activity. J. Biol. Chem. 281, 31963–31971.
Miot, M., Betton, J.M., 2004. Protein quality control in the bacterial periplasm. Microb. Cell Fact. 3, 4.
Missiakas, D., Betton, J.M., Raina, S., 1996. New components of protein folding in extracytoplasmic compartments of *Escherichia coli* SurA, FkpA and Skp/OmpH. Mol. Microbiol. 21, 871–884.
Mücke, M., Schmid, F.X., 1994. Intact disulfide bonds decelerate the folding of ribonuclease T1. J. Mol. Biol. 239, 713–725.
Narayanan, T., Diat, O., Bösecke, P., 2001. SAXS and USAXS on the high brillance beamline at the ESRF. Nucl. Instrum. Methods Phys. Res. Sect. A 467–468, 1005–1009.
Oostenbrink, C., Soares, T.A., van der Vegt, N.F., van Gunsteren, W.F., 2005. Validation of the 53A6 GROMOS force field. Eur. Biophys. J. 34, 273–284.
Petoukhov, M.V., Svergun, D.I., 2005. Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys. J. 89, 1237–1250.
Putnam, C.D., Hammel, M., Hura, G.L., Tainer, J.A., 2007. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q. Rev. Biophys. 40, 191–285.
Rahfeld, J.U., Schierhorn, A., Mann, K., Fischer, G., 1994. A novel peptidyl-prolyl *cis/trans* isomerase from *Escherichia coli*. FEBS Lett. 343, 65–69.
Ranganathan, R., Lu, K.P., Hunter, T., Noel, J.P., 1997. Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. Cell 89, 875–886.
Riboldi-Tunnicliffe, A., König, B., Jessen, S., Weiss, M.S., Rahfeld, J., Hacker, J., Fischer, G., Hilgenfeld, R., 2001. Crystal structure of Mip, a prolylisomerase from *Legionella pneumophila*. Nat. Struct. Biol. 8, 779–783.
Rouviere, P.E., Gross, C.A., 1996. SurA, a periplasmic protein with peptidyl-prolyl isomerase activity, participates in the assembly of outer membrane porins. Genes Dev. 10, 3170–3182.
Saul, F.A., Arié, J.P., Vulliez-le Normand, B., Kahn, R., Betton, J.M., Bentley, G.A., 2004. Structural and functional studies of FkpA from *Escherichia coli*, a *cis/trans* peptidyl-prolyl isomerase with chaperone activity. J. Mol. Biol. 335, 595–608.
Scholz, C., Rahfeld, J., Fischer, G., Schmid, F.X., 1997a. Catalysis of protein folding by parvulin. J. Mol. Biol. 273, 752–762.
Scholz, C., Stoller, G., Zarnt, T., Fischer, G., Schmid, F.X., 1997b. Cooperation of enzymatic and chaperone functions of trigger factor in the catalysis of protein folding. EMBO J. 16, 54–58.
Scholz, C., Mücke, M., Rape, M., Pecht, A., Pahl, A., Bang, H., Schmid, F.X., 1998. Recognition of protein substrates by the prolyl isomerase trigger factor is independent of proline residues. J. Mol. Biol. 277, 723–732.
Schulze-Gahmen, U., Aono, S., Chen, S., Yokota, H., Kim, R., Kim, S.H., 2005. Structure of the hypothetical *Mycoplasma* protein MPN555 suggests a chaperone function. Acta Crystallogr. D Biol. Crystallogr. 61, 1343–1347.
Sekerina, E., Rahfeld, J.U., Muller, J., Fanghanel, J., Rascher, C., Fischer, G., Bayer, P., 2000. NMR solution structure of hPar14 reveals similarity to the peptidyl prolyl *cis/trans* isomerase domain of the mitotic regulator hPin1 but indicates a different functionality of the protein. J. Mol. Biol. 301, 1003–1017.
Semenyuk, A.V., Svergun, D., 1991. GNOM – a program package for small-angle scattering data processing. J. Appl. Crystallogr. 24, 537–540.
Sklar, J.G., Wu, T., Kahne, D., Silhavy, T.J., 2007. Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in *Escherichia coli*. Genes Dev. 21, 2473–2484.
Stirling, P.C., Bakhoum, S.F., Feigl, A.B., Leroux, M.R., 2006. Convergent evolution of clamp-like binding sites in diverse chaperones. Nat. Struct. Mol. Biol. 13, 865–870.
Svergun, D.I., Koch, M.H.J., 2003. Small-angle scattering studies of biological macromolecules in solution. Rep. Prog. Phys. 66, 1735–1782.
Svergun, D., Barberato, C., Koch, M.H., 1995. CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J. Appl. Crystallogr. 28, 768–773.
Terada, T., Shirouzu, M., Fukumori, Y., Fujimori, F., Ito, Y., Kigawa, T., Yokoyama, S., Uchida, T., 2001. Solution structure of the human parvulin-like peptidyl prolyl *cis/trans* isomerase, hPar14. J. Mol. Biol. 305, 917–926.
Tossavainen, H., Permi, P., Purhonen, S.L., Sarvas, M., Kilpelainen, I., Seppala, R., 2006. NMR solution structure and characterization of substrate binding site of the PPIase domain of PrsA protein from *Bacillus subtilis*. FEBS Lett. 580, 1822–1826.
Trapani, S., Navaza, J., 2008. AMoRe: classical and modern. Acta Crystallogr. D Biol. Crystallogr. 64, 11–16.
Tsutakawa, S.E., Hura, G.L., Frankel, K.A., Cooper, P.K., Tainer, J.A., 2007. Structural analysis of flexible proteins in solution by small angle X-ray scattering combined with crystallography. J. Struct. Biol. 158, 214–223.
Uson, I., Sheldrick, G.M., 1999. Advances in direct methods for protein crystallography. Curr. Opin. Struct. Biol. 9, 643–648.
Volkov, V.V., Svergun, D.I., 2003. Uniqueness of ab initio shape determination in small-angle scattering. J. Appl. Crystallogr. 36, 860–864.
Wall, M.E., Ealick, S.E., Gruner, S.M., 1997. Three-dimensional diffuse x-ray scattering from crystals of Staphylococcal nuclease. Proc. Natl. Acad. Sci. USA 94, 6180–6184.
Wilson, M.A., Brunger, A.T., 2003. Domain flexibility in the 1.75 A resolution structure of Pb2+-calmodulin. Acta Crystallogr. D Biol. Crystallogr. 59, 1782–1792.
Wohlkönig, A., Hodak, H., Clantin, B., Senechal, M., Bompard, C., Jacob-Dubuisson, F., Villeret, V., 2008. Crystallization and preliminary X-ray diffraction analysis of the peptidylprolyl isomerase Par27 of *Bordetella pertussis*. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 64, 809–812.
Wu, S., Zhang, Y., 2007. LOMETS: a local meta-threading-server for protein structure prediction. Nucleic Acids Res. 35, 3375–3382.
Wülfing, C., Plückthun, A., 1994. Protein folding in the periplasm of *Escherichia coli*. Mol. Microbiol. 12, 685–692.
Wyatt, P.J., 1998. Submicrometer particle sizing by multiangle light scattering following fractionation. J. Colloid Interface Sci. 197, 9–20.
Xu, X., Wang, S., Hu, Y.X., McKay, D.B., 2007. The periplasmic bacterial molecular chaperone SurA adapts its structure to bind peptides in different conformations to assert a sequence preference for aromatic residues. J. Mol. Biol. 373, 367–381.

III. *Arrière de couverture du livre « XIV International Conference on Negative Strand Viruses », Bruges, Belgique, 20-25 juin 2010.*



Nanobiological objects built with Negative-Strand RNA virus proteins.

A synthetic peptide corresponding to the N-terminal sub-domain of rabies virus nucleoprotein including an engineered disulfide bridge forms amyloid fibrils (upper picture). The ribbon representation in the inset shows an excerpt of the crystal structure of the recombinant circular nucleoprotein-RNA complex (2gtt.pdb) corresponding to this sub-domain. When recombinant N-RNA rings are added to the fibril solution, these bind to the fibrils in an oriented manner and completely cover their surface (lower picture). The ribbon representations show the crystal structure of circular N-RNA complexes. The bottom pictures show different examples of these nanobiological objects.

Cédric Leyrat, Francine Gérard, Irina Gustche and Marc Jamin, UVHCI UMI3265 UJF-EMBL-CNRS, Grenoble, France